COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 35 (2011) 1518–1552

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/j.1551-6709.2011.01197.x

A Quantum Probability Account of Order Effects in Inference

Jennifer S. Trueblood, Jerome R. Busemeyer

Cognitive Science Program, Indiana University

Received 30 September 2010; received in revised form 10 February 2011; accepted 12 February 2011

Abstract

Order of information plays a crucial role in the process of updating beliefs across time. In fact, the presence of order effects makes a classical or Bayesian approach to inference difficult. As a result, the existing models of inference, such as the belief-adjustment model, merely provide an ad hoc explanation for these effects. We postulate a quantum inference model for order effects based on the axiomatic principles of quantum probability theory. The quantum inference model explains order effects by transforming a state vector with different sequences of operators for different orderings of information. We demonstrate this process by fitting the quantum model to data collected in a medical diagnostic task and a jury decision-making task. To further test the quantum inference model, a new jury decision-making experiment is developed. Using the results of this experiment, we compare the quantum inference model with two versions of the belief-adjustment model, the adding model and the averaging model. We show that both the quantum model and the adding model provide good fits to the data. To distinguish the quantum model from the adding model, we develop a new experiment involving extreme evidence. The results from this new experiment suggest that the adding model faces limitations when accounting for tasks involving extreme evidence, whereas the quantum inference model does not. Ultimately, we argue that the quantum model provides a more coherent account for order effects that was not possible before.

Keywords: Inference; Jury decision-making; Recency effects; Belief-adjustment model; Quantum probability theory

1. Introduction

One of the oldest and most reliable findings regarding human inference is that the order of evidence affects the final judgment (Hogarth & Einhorn, 1992). Order effects arise in a number of different inference tasks ranging from judging the likelihood of selecting balls

Correspondence should be sent to Jennifer S. Trueblood, Indiana University Cognitive Science Program, 819 Eigenmann, 1910 E. 10th St. Bloomington, IN 47406. E-mail: cogscij@indiana.edu

from urns (Shanteau, 1970) to judging the guilt or innocence of a defendant in a mock trial (Furnham, 1986; Walker, Thibaut, & Andreoli, 1972). Specifically, we define an order effect to occur when judgments about the probability of a hypothesis given a sequence of information, say A followed by B, does not equal the probability of the same hypothesis when the given information is reversed, B followed by A. Classical probability theory requires $\Pr(X \cap Y \mid H) = \Pr(Y \cap X \mid H)$ for events X and Y, which according to Bayes rule, implies $\Pr(H \mid A \cap B) = \Pr(H \mid B \cap A)$, and so a simple Bayesian model has difficulty accounting for order effects. We will discuss these difficulties in more detail in a later section. As simple probabilistic accounts of order effects are problematic, alternative models of inference have been proposed, such as the averaging model (Shanteau, 1970) and the belief-adjustment model (Hogarth & Einhorn, 1992). These heuristic models lack an axiomatic foundation and only provide an ad hoc explanation for order effects. This article presents a more coherent account of order effects derived from the axiomatic principles of quantum probability theory.

At first, it might seem odd to apply quantum theory to cognitive phenomena. Before we address this general issue, we point out that we are not claiming the brain to be a quantum computer; rather we only use quantum principles to derive cognitive models and leave the neural basis for later research. That is, we use the mathematical formalism of quantum theory without attaching the physical meaning associated with quantum physics. This approach is similar to the application of complexity theory or stochastic processes to domains outside of physics.[1]

There are four reasons for considering a quantum approach to human judgments: (a) judgment is not a simple readout from a preexisting or recorded state, but instead it is constructed from the current context and question; from this first point it then follows that (b) making a judgment changes the context which disturbs the cognitive system; and the second point implies that (c) changes in context produced by the first judgment affects the next judgment producing order effects, so that (d) human judgments do not obey the commutative rule of classic probability theory. If we replace ''human judgment'' with ''physical measurement'' and replace ''cognitive system'' with ''physical system,'' then these are the same points faced by physicists in the 1920s that forced them to develop quantum theory. In other words, quantum theory was initially invented to explain findings in physics that seemed paradoxical from a classical point of view. Similarly, paradoxical findings in cognitive psychology, such as order effects in human judgments, suggest that classical probability theory might be too limited to fully explain various aspects of human cognition. Other paradoxical observations in human judgment and decision making include violations of the sure thing axiom of decision making (Tversky & Shafir, 1992) and violations of the conjunctive and disjunctive rules of classic probability theory (Gilovich, Griffin, & Kahneman, 2002). A growing number of researchers are exploring the use of quantum probability theory to account for these phenomena. For example, a quantum model was developed to account for violations of the sure thing principle (Pothos & Busemeyer, 2009) as well as conjunction and disjunction errors (Franco, 2009; Busemeyer, Pothos, Franco, & Trueblood, 2011). Quantum probability has also been used to model incompatibility and interference effects that arise in other areas of human judgments (Aerts & Aerts, 1994; Khrennikov, 2004).

1.1. Classifying inference tasks

Order effects vary with different inference tasks. Two of the most commonly studied order effects are recency effects, resulting from disproportionate importance of recent evidence, and primacy effects, resulting from disproportionate importance of initial evidence. Fig. 1 illustrates these two effects when two pieces of information are processed. Hogarth and Einhorn (1992) examined order effects, specifically primacy and recency, by classifying the results of 60 experiments according to complexity, length of series, and response mode. The complexity of a task, either complex or simple, was judged by the amount of information subjects were required to process. For example, if evidence items contained more than 600 words, they were classified as complex. The length of the series referred to the number of pieces of evidence that needed to be processed. Hogarth and Einhorn classified tasks as either short, with 2–12 items being processed, or long, with more than 12 items being processed. The response mode was either Step-by-Step, where subjects provided a probability judgment after each piece of evidence, or End-of-Sequence, where subjects only provided a probability judgment after all evidence had been processed.

In this article, we examine simple, Step-by-Step tasks with short series of evidence. As Hogarth and Einhorn discovered that recency effects are prevalent in these types of tasks, we begin by giving a quantum probability account of recency effects in a medical decision-making task. Later, we show that the quantum inference model can also account for data

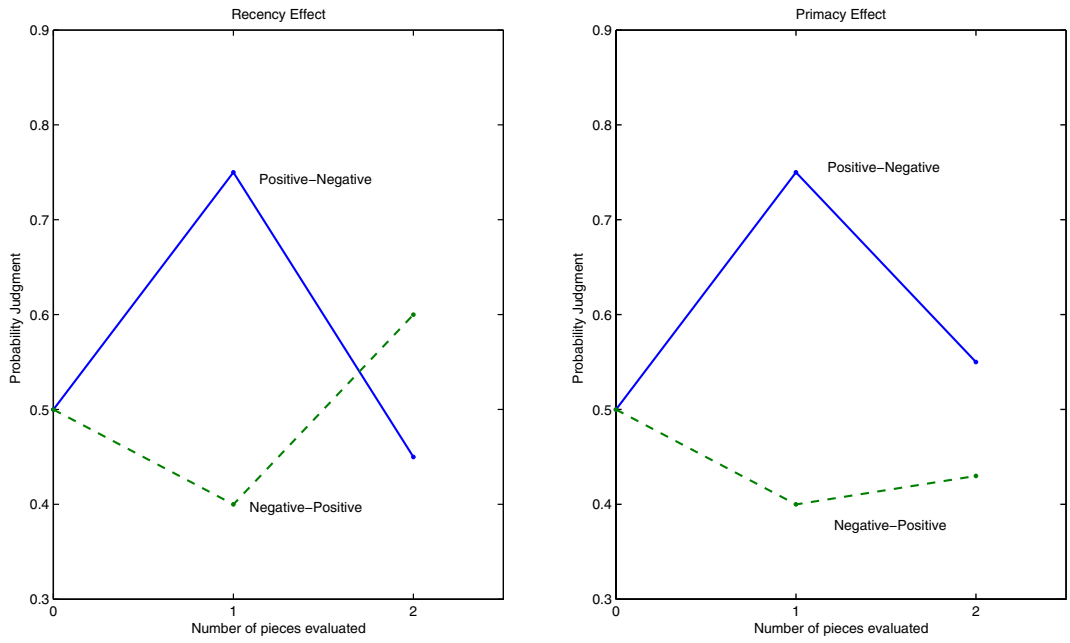

Fig. 1. Primacy (right) and recency (left) order effects. Recency effects have a notable crossover in the two curves associated with the two orderings. Primacy effects do not demonstrate such a crossover.

collected in tasks involving both manipulations of order and evidence strength. Specifically, we examine a jury decision-making task by McKenzie, Lee, and Chen (2002) in which individuals judge the probability of a defendant's guilt given a prosecution and defense of different strengths. Then, we discuss several new experiments examining the effects of order and evidence strength on probability judgments. To aid in the presentation of the quantum inference model, we first provide a description of the medical decision-making task and give a tutorial on the fundamentals of quantum probability through illustrative examples.

1.2. A medical inference task

Bergus, Chapman, Levy, Ely, and Oppliger (1998) examined how the order of clinical data can influence decisions in diagnostic tasks. In this study, 315 active members of the Iowa Academy of Family Practice were asked to estimate the probability that a specific patient had a urinary tract infection (UTI) given the patient's history and physical examination along with laboratory data. The physicians were divided into two groups with one group receiving the history and physical examination information first (H&P-first) and the other receiving the laboratory data first (H&P-last). The history and physical examination were designed to suggest a UTI, whereas the laboratory data were inclusive about the existence of a UTI.

Before either group received any specific information, they were informed that the patient was a 28-year-old woman complaining about ''discomfort on urination.'' After reading this chief complaint, all of the physicians estimated the probability of a UTI. For this initial probability judgment, the H&P-first and the H&P-last groups had almost identical mean estimates of a UTI (67.4% vs. 67.8% , $p = .85$). After the physicians read the first piece of information, the H&P-first group increased their probability estimate of a UTI and the H&P-last group decreased their probability estimate of a UTI (77.8% vs. 44.0% , $p < .001$). This should not come as a surprise because the history and physical examination provided evidence of a UTI and the laboratory data did not. In the final probability estimate, the H&P-first group judged the probability of a UTI to be significantly less than the H&P-last group (50.9% vs. 59.1% , $p = .03$). The data suggest that physicians placed greater salience on the final piece of information resulting in a recency effect. The results are summarized in Table 1.

One might argue that the order effects observed in this study are simply caused by forgetting or memory failures. This seems unlikely because, for the medical decision-making task, participants read short summaries of information without any time delays between each presentation of the information. Furthermore, there is evidence that sequential judgments such as these are made on-line by a sequential updating procedure and are not memory-based

Table 1
Mean probability estimates from diagnostic task

	H&P-first	H&P-last
Initial	Pr(UTI) = .674	Pr(UTI) = .678
Second	Pr(UTI | H&P) = .778	Pr(UTI | Lab) = .440
Final	Pr(UTI | H&P, Lab) = .509	Pr(UTI | Lab, H&P) = .591

(Hastie & Park, 1986). In fact, the correlation between judgments and recall of information is found to be negligible (Anderson & Hubert, 1963). In sum, the order effects are the result of the sequential updating procedure used to revise beliefs rather than memory failures.

1.3. Problems for a simple Bayesian model

A Bayesian inference model has difficulty accounting for order effects. To account for order effects, a naive Bayesian model must introduce presentation order as another piece of information. Suppose we are interested in evaluating the probability of a hypothesis, H, given two pieces of information, A and B. As classical probability obeys the commutative property, we have

$$p(H \mid A \cap B) = p(H \mid A) \cdot \frac{p(B \mid H \cap A)}{p(B \mid A)} = p(H \mid B) \cdot \frac{p(A \mid H \cap B)}{p(A \mid B)} = p(H \mid B \cap A).$$

Thus, to model order effects, the Bayesian model needs two more events: the event O_1 that A is presented before B and event O_2 that B is presented before A. In this case, we obtain $p(H \mid A \cap B \cap O_1) \neq p(H \mid A \cap B \cap O_2)$. However, without specifying $p(H) \times p(O_i \mid H) \times p(A \mid H \cap O_i) \times p(B \mid H \cap O_i \cap A)$, this approach simply redescribes empirical results, and such a specification is not known at present. In many empirical studies of order effects, the presentation order is randomly determined. Thus, the major difficulty for the Bayesian model is that order information is often irrelevant.

2. Quantum inference model

2.1. State vector

This section develops a quantum model using the medical inference as an example. Later, we extend the model for the jury inference task. For both applications, our model assumes there are two complementary hypotheses, h_1 and h_2. We also postulate the existence of positive evidence, e_1, for h_1 and negative evidence, e_2, for h_1. Because h_1 and h_2 are complementary, e_2 can also be interpreted as positive evidence for h_2. We assume the positive and negative evidence comes from two different sources of information, A and B. For the medical decision-making task, we let $h_1 =$ UTI present and $h_2 =$ UTI absent. Information source A corresponds to the history and physical examination, and information source B corresponds to the laboratory data. Note that the history and physical examination, A, provides positive evidence, e_1, for the presence of a UTI, h_1. In contrast, the laboratory data, B, provides evidence, e_2, for the absence of a UTI, h_2.

Our quantum model represents an individual's initial state of belief by a state vector denoted $|\psi\rangle$. [2] The belief state $|\psi\rangle$ can be expressed as a linear combination or superposition of four basis states $\{|N_{11}\rangle, |N_{12}\rangle, |N_{21}\rangle, |N_{22}\rangle\}$, where each basis state $|N_{ij}\rangle$ is a vector with all zeros except a one located in the row corresponding to pattern $h_i \wedge e_j$. For example, $|N_{12}\rangle$ is a belief state in which the probability equals one for $h_1 \wedge e_2$, in which case the person is

certain that h_1 is true and negative evidence, e_2, is present. With respect to these basis vectors, the general belief state can be expressed as a superposition (i.e., a linear combination)

$$|\psi\rangle = \sum \omega_{ij} \cdot |N_{ij}\rangle.$$

In this way, we see that the belief state is a vector within a four-dimensional vector space spanned by the $N = \{|N_{11}\rangle, |N_{12}\rangle, |N_{21}\rangle, |N_{22}\rangle\}$ basis. In this formulation, a probability amplitude ω_{ij} determines the belief about a pattern of features, $h_i \wedge e_j$. Specifically, for the medical decision-making task, the belief state vector can be represented by the four amplitudes:

$$\omega = \begin{bmatrix} \omega_{h_1,e_1} \\ \omega_{h_1,e_2} \\ \omega_{h_2,e_1} \\ \omega_{h_2,e_2} \end{bmatrix} = \begin{bmatrix} \text{amplitude for UTI present and positive evidence} \\ \text{amplitude for UTI present and negative evidence} \\ \text{amplitude for UTI absent and positive evidence} \\ \text{amplitude for UTI absent and negative evidence} \end{bmatrix}. \quad (1)$$

We define the amplitudes ω_{ij} such that $|\psi\rangle$ is a unit length vector. The quantum probability of a particular feature pattern is $q(h_i \wedge e_j) = ||\omega_{ij}||^2$. In other words, the probability of inferring the feature pattern $h_i \wedge e_j$ assuming the belief state, $|\psi\rangle$, is $\|\omega_{ij}\|^2$. For example, in the medical decision-making task, the probability of a UTI infection and positive evidence is $q(h_1 \wedge e_1) = ||\omega_{11}||^2$. Note that the probabilities, $q(h_i \wedge e_j)$, sum to one across all four feature patterns.

2.2. Projectors

In quantum theory, probabilities are computed by first projecting the belief state onto a subspace that defines an event, and then computing the squared modulus of the projection. Here, we define projectors that project unit vectors from our vector space onto individual basis vectors, or feature patterns. For example, the belief state is projected onto the event $h_1 \wedge e_1$ by the projector:

$$P_{11} = P(h_1, e_1) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

If we are interested in calculating the probability of the joint event the UTI is present and evidence 1 is present, we project our state vector $|\psi\rangle$ onto the basis vector $|N_{11}\rangle$:

$$P(h_1, e_1)|\psi\rangle = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \omega_{11} \\ \omega_{12} \\ \omega_{21} \\ \omega_{22} \end{bmatrix} = \begin{bmatrix} \omega_{11} \\ 0 \\ 0 \\ 0 \end{bmatrix}.$$

Then, we calculate the quantum probability of the event the UTI is present and evidence 1 is present by the squared length of the projection, $\|P(h_1, e_1)|\psi\rangle\|^2 = \|\omega_{11}\|^2$.

Suppose we are interested in the probability of the event h_1 = ''UTI is present.'' This event contains the two feature patterns $\{h_1 \wedge e_1, h_1 \wedge e_2\}$. In quantum theory, events are represented as subspaces of our vector space, and the event h_1 is represented by the span of the basis vectors $\{|N_{11}\rangle, |N_{12}\rangle\}$ which corresponds to the projector $P(h_1) = P_{11} + P_{12}$. For example, the event UTI is present is represented by the projector:

$$P(h_1) = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

To find the probability of this event, we calculate

$$q(h_1) = ||P(h_1)|\psi\rangle||^2 = ||\omega_{11}||^2 + ||\omega_{12}||^2.$$

2.3. State revision

After observing an event, the quantum state undergoes revision. Suppose we start with state $|\psi\rangle$, and we learn that the UTI is in fact present. Then, we project $|\psi\rangle$ onto the subspace corresponding to this event and normalize this projection (so that the revised state remains unit length):

$$|\psi_{h_1}\rangle = \frac{P(h_1)|\psi\rangle}{||P(h_1)|\psi\rangle\ ||}.$$

Normalization ensures that the length of this new state vector, $|\psi_{h_1}\rangle$ equals one. Psychologically, state revision corresponds to updating the belief state after processing new information.

2.4. Compatibility

The concept of compatibility is one of the most important new ideas introduced by quantum theory. The essential idea is that a vector space representation of beliefs allows one to change the basis vectors used to describe the belief state. In other words, we are not restricted to describing the belief state in terms of the feature patterns represented by the basis vectors $\{|N_{11}\rangle, |N_{12}\rangle, |N_{21}\rangle, |N_{22}\rangle\}$. When considering a new source of information, we can change our perspective by rotating to a new set of basis vectors that can be used to represent the beliefs from the point of view of the new source.

To see how this works, we will draw upon a two-dimensional example to illustrate the concept of compatibility because it is impossible to visualize our four-dimensional vector space. Suppose we are interested in representing two events, truth and goodness. Let us assume that truth is defined in terms of two basis vectors $|x_1\rangle$ corresponding to the truth

value ''true'' and $|x_2\rangle$ corresponding to the truth value ''false.'' We also assume that goodness is represented within the same two-dimensional vector space but is defined in terms of two different basis vectors, $|y_1\rangle$ corresponding to ''good'' and $|y_2\rangle$ corresponding to ''bad.'' The vector S represents the person's belief state concerning a statement. To determine the probability that a statement is ''true,'' the belief state S is projected on the $|x_1\rangle$ basis vector (and the projection on $|x_2\rangle$ determines the probability of ''false''). But to determine the probability that the statement is ''good,'' the belief state S is projected on the $|y_1\rangle$ basis vector (and the projection on the $|y_2\rangle$ basis vector determines the probability of ''bad''). In other words, our belief state S can be represented by $\{|x_1\rangle, |x_2\rangle\}$ or $\{|y_1\rangle, |y_2\rangle\}$ where the $|y_j\rangle$ basis is achieved by ''rotating'' the $|x_j\rangle$ basis. So our vector space is represented by more than one set of basis vectors. In this example, we are assuming that truth and goodness are *incompatible* events—a statement cannot be exactly true and exactly good at the same time. In other words, if you are sure something is true, then you are uncertain whether or not it is good, and vice versa (see Fig. 2 for an illustration of this example). From a cognitive perspective, different feature patterns are needed to describe different events. If events can be represented by using the same set of feature patterns, so that there is no need to change the basis, then we call the events compatible.

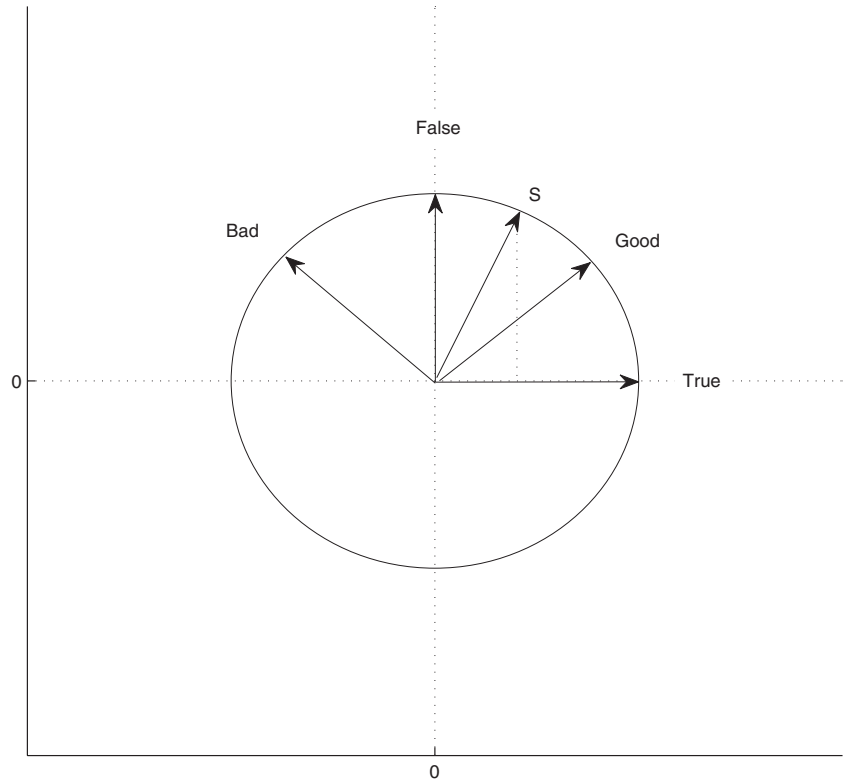

Fig. 2. Truth and goodness are represented as incompatible events. The belief state S is being projected onto the subspace corresponding to the truth value ''true.''

In the case when two events are incompatible, we apply a unitary transformation to ''rotate'' one basis to the other. A unitary transformation is a matrix U that satisfies $U^{\dagger}U = I$, where I is the identity matrix and $U^{\dagger}$ is the conjugate transpose of U. The matrix U must be unitary to preserve lengths (we wish the state vector and the basis vectors to have length one) and to preserve inner products (we wish the basis vectors to remain orthogonal). For example, we might postulate the basis vectors for goodness are $\{|y_1\rangle, |y_2\rangle\} = \{U\cdot|x_1\rangle, U\cdot|x_2\rangle\}$, where U is the rotation matrix given by

$$U = \begin{bmatrix} \cos\frac{\pi}{4} & -\sin\frac{\pi}{4} \\ \sin\frac{\pi}{4} & \cos\frac{\pi}{4} \end{bmatrix} = \begin{bmatrix} 0.7071 & -0.7071 \\ 0.7071 & 0.7071 \end{bmatrix}.$$

As U is unitary, we must have the basis vectors for truth be defined as $\{|x_1\rangle, |x_2\rangle\} = \{U^{\dagger}\cdot|y_1\rangle, U^{\dagger}\cdot|y_2\rangle\}$. In general, a state vector can be represented with respect to either basis. If $|\psi\rangle$ is the state vector described with respect to the $|x_j\rangle$ basis, then $U\cdot|\psi\rangle$ is the same state vector described with respect to the $|y_j\rangle$ basis. Psychologically, a person has a state of beliefs, but this same state of beliefs can be evaluated from different perspectives or viewpoints.

In summary, quantum probability is a theory that describes how to assign probabilities to events, but unlike classical probability, quantum probability allows for events to be incompatible. This important feature of quantum probability allows us to model phenomena such as order effects.

2.5. Representation of different sources of information

To provide a point of comparison, we begin by describing a classical probability formulation of the problem. In a classical probability model, we might define our sample space S in terms of the two hypotheses h_1 and h_2 and the three observations N, A, and B, where N represents the initial description of the task. We might further assume that each observation has two possible outcomes, e_1 and e_2. These outcomes represent the type of evidence provided by the observations. We consider the initial description of the problem as providing evidence because it typically contains some information even if this information is not particularly strong. For example, in the medical inference task, subjects were initially informed that the patient was a woman complaining about ''discomfort on urination.'' Even this brief description provides subjects with some evidence about the hypotheses.

The sample space S represents all possible combinations of hypotheses and observations. Because there are two hypotheses and three observations each with two outcomes, the dimension of the sample space is $n = 2^4 = 16$. Elementary elements of S have the form $h_i \wedge (N = e_j) \wedge (A = e_k) \wedge (B = e_l)$, and it is possible to construct a direct correspondence between these elementary events and the basis vectors of a vector space. For simplicity, let us index the elementary events such that $z_1 = h_1 \wedge (N = e_1) \wedge (A = e_1) \wedge (B = e_1)$ and $z_n = h_2 \wedge (N = e_2) \wedge (A = e_2) \wedge (B = e_2)$. With the sample space defined as $\{z_1,\ldots, z_n\}$, we say that the corresponding vector space has basis $\{|z_1\rangle,\ldots, |z_n\rangle\}$. When using this basis to represent vectors in the vector space, the $n \times 1$ column vector that has all zeros except for a one in row k is a coordinate representation of the basis vector $|z_k\rangle$.

We associate this basis vector with the elementary event z_k. Thus, we can represent the elementary events of classical probability theory as a basis of a vector space. This vector space representation provides a useful way to compare the classical and quantum probability models.

As we have mentioned, the classical probability model has difficulty accounting for order effects because the commutative property holds. Now, we present our idea of how different sources of information are represented in the quantum model. We assume an individual has different representations for beliefs depending on three different points of view: a point of view determined by the initial description of the task denoted by N, a point of view determined by the presentation of source A, and a point of view determined by the presentation of source B. The quantum inference model accounts for order effects by assuming the three points of view N, A, and B are incompatible. By allowing the three observations to be incompatible, the 16-dimensional vector space needed for the classical model is reduced to a four-dimensional space in the quantum model. Further, each point of view is associated with a different basis for this four-dimensional vector space. We assume that the hypotheses are compatible with the two types of evidence allowing us to define the joint events $h_i \wedge e_j$.

We assume that an individual adopts the three points of view throughout the course of the task. When the individual takes the initial point of view, we use the basis vectors $|N_{ij}\rangle$, where $i = h_1$ or h_2 and $j = e_1$ or e_2. When the individual changes his or her point of view after seeing A, the basis vectors $|A_{ij}\rangle$ are used, and when the individual changes his or her point of view after seeing B, the basis vectors $|B_{ij}\rangle$ are used. For example, in the medical decision-making task, the physicians change their points of view after reading the history and physical information and again after reading the laboratory data. In other words, the physicians are viewing information about the presence of the UTI and evidence from different perspectives or interpretations depending on the source of information.

The same quantum belief state of the individual, $|\psi\rangle$ can be represented with respect to any one of these three points of views:

$$|\psi\rangle = \sum \omega_{ij} \,|\, N_{ij}\rangle = \sum \alpha_{ij} \,|\, A_{ij}\rangle = \sum \beta_{ij} \,|\, B_{ij}\rangle.$$

Each basis (a set of four orthonormal basis vectors) spans the same four-dimensional vector space. The unit column vectors ω, α, and β represent the probability amplitudes with respect to the N basis, the A basis, and the B basis, respectively. For example, within the A and B views, we have

$$\alpha = \begin{bmatrix} \alpha_{h_1,e_1} \\ \alpha_{h_1,e_2} \\ \alpha_{h_2,e_1} \\ \alpha_{h_2,e_2} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_{h_1,e_1} \\ \beta_{h_1,e_2} \\ \beta_{h_2,e_1} \\ \beta_{h_2,e_2} \end{bmatrix}.$$

The probability amplitude α_{h_i,e_j} corresponds to the amplitude that feature pattern h_i is present and evidence e_j is present when considering the A source, and similarly, the probability amplitude β_{h_i,e_j} corresponds to the amplitude that feature pattern h_i is present and evidence e_j is present when considering the B source.

2.6. Unitary transformations

Unitary transformations relate one point of view to another and correspond to an individual's shifts in perspective. For example, if a physician is first presented with the history and physical information followed by the laboratory data, he or she would start with an initial perspective, represented by ω, shift to the history and physical perspective, represented by α, and finally shift to the laboratory perspective, represented by β. In this model, U_{AN} transforms the probability amplitudes from the initial view to the point of view associated with the A basis, and U_{BN} transforms the amplitudes from the initial basis into the B basis. So we have

$$\alpha = U_{AN}\omega, \quad \omega = U_{AN}^{\dagger}\alpha,$$

$$\beta = U_{BN}\omega, \quad \omega = U_{BN}^{\dagger}\beta.$$

The unitary transformation U_{BA} transforms the amplitudes from the A basis into the B basis, and U_{AB} transforms the amplitudes from the B basis into the A basis. In the medical example, the first transformation, U_{BA}, corresponds to a change of perspective after seeing the history and physical followed by the laboratory data. The second transformation, U_{AB}, corresponds to a shift in perspective after seeing the laboratory data followed by the history and physical information. These last two transformations are derived from the two previously described transformation as follows:

$$U_{BA} = U_{BN}U_{AN}^{\dagger}, \quad U_{AB} = U_{AN}U_{BN}^{\dagger}.$$

There is nothing significant about having the transformations pass through the initial point of view. In fact, if we had defined unitary matrices U_{BC} and U_{AC}, where C refers to some other basis of our vector space, then we could have defined U_{BA} and U_{AB} in terms of these matrices instead. This is due to the fact that unitary matrices form a group. Because all of the unitary transformations in this model are defined in terms of U_{AN} and U_{BN}, we will now focus our attention on constructing these two transformations. Readers uninterested in the mathematical derivation of the unitary matrices can skip over the next section and proceed to the section ''Details for the medical example'' without losing the conceptual ideas.

2.7. Construction of U_{AN} and U_{BN}

Any unitary matrix can be constructed from the matrix exponential function $U = e^{-i\theta H}$ of a Hermitian matrix H that satisfies $H^{\dagger} = H$ (Nielsen & Chuang, 2000).[3] The complex number i appearing in the matrix exponential is required to maintain the unitary property. Thus, to construct U_{AN} and U_{BN}, we must first define a Hermitian matrix H and two parameters θ_A and θ_B. Specifically, we define $H = H_1 + H_2$, where H_1 and H_2 are also Hermitian matrices.

To explain the specific form of our Hermitian matrix H, we begin by examining the component matrix H_1. We proceed by first formulating a Hermitian matrix for a simpler two-dimensional problem, and then extending this to define H_1 for the full four-dimensional

problem. Suppose our vector space is defined as the span of the two basis vectors $|h_1 \wedge e_1\rangle$ and $|h_1 \wedge e_2\rangle$. Notice that this is the subspace of our original four-dimensional vector space corresponding to hypothesis h_1. We continue to postulate that decision makers view the space from three perspectives: n, a, and b. (We will use lowercase letters when referring to the two-dimensional problem to distinguish it from the full model.) Thus, we must define U_{an} and U_{bn} to relate the three points of view. These unitary matrices depend upon our construction of a two-dimensional Hermitian matrix W.

To start, any 2×2 Hermitian matrix can be expressed as a linear combination of the Pauli matrices:

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}; \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}; \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Specifically, we write

$$W = \eta_x \cdot \sigma_x + \eta_y \cdot \sigma_y + \eta_z \cdot \sigma_z.$$

Thus, for $j = a$ and b, the corresponding unitary matrices are

$$U_{jn} = e^{-i\theta_j W} = e^{-i\theta_j(\eta_x \cdot \sigma_x + \eta_y \cdot \sigma_y + \eta_z \cdot \sigma_z)}.$$

We also assume that $(\eta_x^2 + \eta_y^2 + \eta_z^2)^{\frac{1}{2}} = 1$. Applying Euler's identity, we can write the unitary matrices as

$$U_{jn} = \cos(\theta_j) \cdot I - i \sin(\theta_j) \cdot (\eta_x \cdot \sigma_x + \eta_y \cdot \sigma_y + \eta_z \cdot \sigma_z),$$

where I is the identity matrix. This produces a rotation of degree θ_j around the axis defined by the unit length vector (η_x,η_y,η_z). (For more details, please see Nielsen & Chuang, 2000, Chapter 4.)

The probability that $h_1 \wedge e_1$ is true after the rotation U_{jn} is periodic in the variable θ_j. If we wish to maintain an overall probability greater than .5 for one type of evidence over the other across time, then we need to set $\eta_y = 0$. Essentially, this ensures that if evidence type e_1 is associated with source A, then e_1 is favored throughout the entire presentation of A. With $\eta_y = 0$, the probability for one type of evidence over the other will be maximized whenever $\eta_x = \eta_z > 0$. Because the vector (η_x,η_y,η_z) has unit length, we must set $\eta_x = \eta_z = \frac{1}{\sqrt{2}}$. Now, we define the Hermitian matrix W as

$$W = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

Thus, the unitary matrices for the two-dimensional problem are defined as

$$U_{jn} = \exp\left\{ -i\theta_j \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \right\}.$$

Returning to the full four-dimensional model, we can now specify the matrix H_1 in terms of the matrix W. We assume that H_1 is defined as the tensor product given by

$$H_1 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes W = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \otimes \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix}.$$

The unitary matrix with H_1 as a generator transforms the amplitudes according to the strength of the current evidence. For example, suppose we are presented with positive evidence, e_1, for h_1. In this case, H_1 results in rotating the probability amplitudes to favor events involving e_1. In other words, the corresponding unitary matrix strengthens the amplitudes corresponding to e_1 and weakens the amplitudes corresponding to e_2. A similar rotation occurs when negative evidence, e_2, is presented. Further, the unitary matrix corresponding to H_1 strengthens and weakens the evidence amplitudes to the greatest extent possible. This is due to the fact that W was designed to maximize the probability of one type of evidence over the other.

Now, we turn to the construction of the second component H_2 of the Hermitian matrix H. As in the case with H_1, we proceed by first defining a Hermitian matrix for a two-dimensional space. In this case, we consider the vector space spanned by basis vectors $|h_1 \wedge e_1\rangle$ and $|h_2 \wedge e_1\rangle$. Notice that this is the subspace of our full four-dimensional vector space corresponding to evidence e_1. At this point, we proceed exactly as we did before. We define a Hermitian matrix V as a linear combination of Pauli matrices. Because we wish to maintain an overall probability greater than .5 for one hypothesis over the other across time, we set $\eta_y = 0$. This ensures that preferences for hypotheses do not reverse during the presentation of a particular source. As before, we set $\eta_x = \eta_z = \frac{1}{\sqrt{2}}$ to maximize the probability of one hypothesis over the other. Thus, we have $V = W$.

Now, if we rearrange the coordinate vector given in Eq. (1) such that

$$\omega = \begin{bmatrix} \omega_{h_1,e_1} \\ \omega_{h_2,e_1} \\ \omega_{h_2,e_2} \\ \omega_{h_1,e_2} \end{bmatrix},$$

we can easily write H_2 in terms of V. Specifically, we have

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

Of course, we want to combine H_1 and H_2. So, we will need to use the same arrangement of coordinates for both matrices. To define H_2 in terms of the coordinates given in Eq. (1), we need to switch some of the rows and columns. We first switch row 2

with row 3 and column 2 with column 3. Then we switch row 2 with row 4 and column 2 with column 4. This gives

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

The Hermitian matrix H_2 results in transforming amplitudes toward h_1 when e_1 is known to be present and toward h_2 when e_2 is known to be present. This matrix evolves an individual's beliefs about the hypotheses and their relationship to the evidence. As in the case of H_1, the unitary matrix corresponding to H_2 evolves the hypothesis amplitudes to the greatest extent possible.

Now, we can define our Hermitian matrix H as

$$H = \frac{1}{\sqrt{2}}(H_1'' + H_2'') = \frac{1}{\sqrt{2}} \left(\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix} \right)$$
$$= \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & 1 & 1 & 0 \\ 1 & -2 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{bmatrix},$$

where H_1'' and H_2'' are the Hermitian matrices corresponding to the unnormalized vector $(\eta_x,\eta_y,\eta_z) = (1,0,1)$. The sum $H_1 + H_2$ coordinates h_1 with e_1 and h_2 with e_2, thereby aligning an individual's belief about the evidence and the hypotheses.

As we have defined our Hermitian matrices, H_1 and H_2, we can now compute U_{AN} and U_{BN} from the matrix exponential functions

$$U_{AN} = e^{\frac{-i}{\sqrt{2}}\theta_A(H_1''+H_2'')}, \quad U_{BN} = e^{\frac{-i}{\sqrt{2}}\theta_B(H_1''+H_2'')}.$$

Note that the model assumes changes in an individual's evaluation of the evidence and belief in the hypotheses occur simultaneously. The eigen decomposition of the unitary matrices and other details are provided in the Appendix.

We have two parameters associated with the unitary transformations. For simplicity, we absorb the normalizing constant $\frac{1}{\sqrt{2}}$ into the parameter values and define $x_A = \frac{1}{\sqrt{2}}\theta_A$ and $x_B = \frac{1}{\sqrt{2}}\theta_B$. Thus, the parameter x_A is needed to define U_{AN} and the other parameter x_B is needed to define U_{BN}. These are the parameters that we will be fitting in the examples below.

One question often raised about quantum models concerns the seemingly large degree of flexibility in these models. To examine this issue, we used the unitary matrix $U_{jN} = e^{-ix_j(H_1'' + H_2'')}$ to evolve beliefs about $h_1 \wedge e_1$, $h_1 \wedge e_2$, $h_2 \wedge e_1$, and $h_2 \wedge e_2$ for 5,000 different values of x_j. We then calculated and plotted the probability of these feature patterns for each parameter value. Fig. 3 shows the probabilities attainable from the quantum model in black. Because quantum probability calculations require the specification of an initial belief state, we assumed that

$$|\psi\rangle = \frac{1}{\sqrt{4}} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

This is equivalent to assuming a uniform distribution over the four feature patterns. The gray area shows all possible probabilities constrained by $\Pr(h_1 \wedge e_1) + \Pr(h_1 \wedge e_2) + \Pr(h_2 \wedge e_1) + \Pr(h_2 \wedge e_2) = 1$. Because of the unitary property of the transformation U_{jN}, all of the calculations from the quantum model satisfy this sum. Thus, we see that probabilities calculated from the quantum model are constrained to specific regions within the larger region of all possible values. We take this as evidence that the quantum model is not overly flexible and cannot account for all possible probability values.

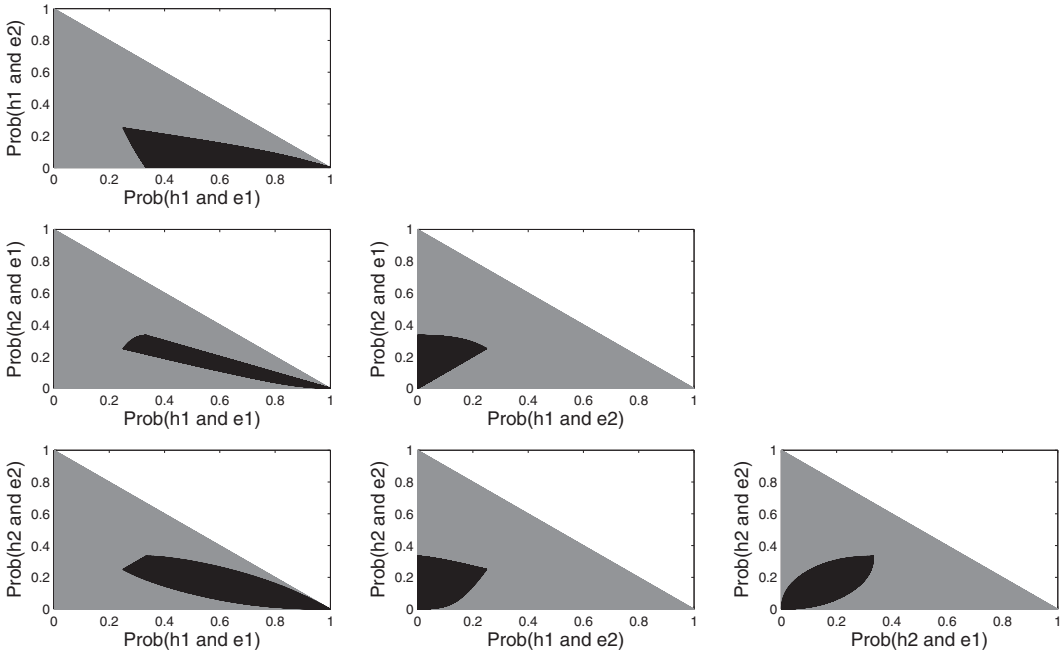

Fig. 3. Probabilities calculated from the quantum model for 5,000 different parameter values are shown in black. All possible probability values are shown in gray.

Further, not any quantum probability model can account for the data. For example, if we redefine the unitary matrices as

$$U_{AN} = e^{-i \cdot x_A \cdot H_1''} \cdot e^{-i \cdot x_A \cdot H_2''}, U_{BN} = e^{-i \cdot x_B \cdot H_1''} \cdot e^{-i \cdot x_B \cdot H_2''},$$

we are unable to fit this model to the order effects found in the medical decision-making task. Note that this alternative quantum model performs sequential transformations of beliefs about evidence and hypotheses during the processing of each source of information. This differs from the first quantum inference model which used unitary transformations that coordinate the hypothesis and evidence simultaneously. (The matrices H_1, H_2 do not commute so that $e^{-i \cdot x_A \cdot H_1''} \cdot e^{-i \cdot x_A \cdot H_2''} \neq e^{-i \cdot x_A \cdot (H_1'' + H_2'')}$). Thus, we find it is necessary for changes in an individual's evaluation of the evidence and belief in the hypotheses to occur in parallel rather than serial.

3. Details for the medical example

To illustrate the quantum inference model's ability to account for recency effects, we fit our model to data collected in the medical decision-making task by Bergus et al. (1998). All of the calculations below were performed with the best-fitting unitary transformations (minimizing sum-of-squared error [SSE]) with parameters $x_A = 4.4045$ and $x_B = 0.3306$. The first parameter x_A coordinates the physician's belief in the presence of the UTI with the history and physical information. The second parameter x_B coordinates the physician's belief in the absence of the infection with the laboratory data. The results from the quantum model are shown in Table 2.

3.1. Inference calculations

During the medical decision-making task, one group of physicians judged the probability that a patient has a UTI (a) before any evidence is presented, (b) after reviewing the patient's history and physical exam, and (c) after reviewing the laboratory work. A second group of physicians also judged the probability of a UTI but reviewed the laboratory work before the history and physical exam. Both groups of physicians initially read a brief description of the patient (a 28-year-old woman) and the chief complaint (discomfort on urination). Like the Bayesian inference model, we must specify the quantum model's priors before we can proceed with the inference calculations. Both order groups were given some

Table 2
Quantum inference model results for diagnostic task

	H &P-first	H &P-last
Initial	Pr(D) = .676	Pr(D) = .676
Second	Pr(D | H &P) = .793	Pr(D | Lab) = .438
Final	Pr(D | H &P, Lab) = .504	Pr(D | Lab, H &P) = .590
SSE	.00025	

initial complaint information before any evidence was presented. Because both the history and physical first group and the history and physical last group gave similar initial probability judgments for a UTI with a mean of .676, we define the initial unit state vector as

$$\omega = \begin{bmatrix} \omega_{h_1,e_1} \\ \omega_{h_1,e_2} \\ \omega_{h_2,e_1} \\ \omega_{h_2,e_2} \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{.676}{2}} \\ \sqrt{\frac{.676}{2}} \\ \sqrt{\frac{.324}{2}} \\ \sqrt{\frac{.324}{2}} \end{bmatrix}.$$

From a cognitive processing point of view, this initial state vector can be thought of as a belief state formed by prior knowledge and the information presented in the initial brief description. This memory state represents the potential for a specific pattern to be retrieved. When physicians are questioned about the probability of a UTI, the initial belief state is projected onto the disease present subspace.

$$P(h_1)\omega = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \sqrt{\frac{.676}{2}} \\ \sqrt{\frac{.676}{2}} \\ \sqrt{\frac{.324}{2}} \\ \sqrt{\frac{.324}{2}} \end{bmatrix} = \begin{bmatrix} \sqrt{\frac{.676}{2}} \\ \sqrt{\frac{.676}{2}} \\ 0 \\ 0 \end{bmatrix}.$$

This projection determines the extent to which the retrieval cue h_1 matches the initial belief state ω (viewed from the initial point of view). The probability of the UTI from the initial perspective equals the squared projection:

$$q(\text{UTI}) = ||P(h_1)\omega||^2 = ||\omega_{h_1,e_1}||^2 + ||\omega_{h_1,e_2}||^2 = .676.$$

Now, we will focus on the group of physicians presented with the history and physical first. (The calculations for the history and physical last group are similar.) We assume that the initial event and the history and physical event cannot be described by a common set of basis vectors. In other words, the two events are incompatible. So we must perform a change of basis by applying the U_{AN} unitary transformation to the initial state:

$$\alpha = U_{AN}\omega.$$

The unit vector α represents the belief state following presentation of the history and physical information. The corresponding vector of squared amplitudes is

$$\alpha \Rightarrow \begin{bmatrix} ||\alpha_{11}||^2 \\ ||\alpha_{12}||^2 \\ ||\alpha_{21}||^2 \\ ||\alpha_{22}||^2 \end{bmatrix} = \begin{bmatrix} .6621 \\ .0752 \\ .1729 \\ .0898 \end{bmatrix}.$$

In the following calculations, we will only show the vectors of squared amplitudes because these quantities are the most interpretable. The history and physical information provides evidence favoring the presence of the UTI, so we project α onto the ''positive evidence'' or e_1 subspace

$$P(\text{h\&p})\alpha = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \alpha \Rightarrow \begin{bmatrix} .6621 \\ 0 \\ .1729 \\ 0 \end{bmatrix}.$$

At this point, the revised quantum state of the physician after reviewing the history and physical information is

$$\alpha_{\text{h\&p}} = \frac{P(\text{h\&p})\alpha}{||P(\text{h\&p})\alpha||} \Rightarrow \begin{bmatrix} .7929 \\ 0 \\ .2071 \\ 0 \end{bmatrix}.$$

Now, when physicians are questioned about the probability of a UTI, the revised state $\alpha_{h\&p}$ is projected onto the disease present subspace. Similar to the initial case, this projection determines how well the revised beliefs $\alpha_{\text{h\&p}}$ match the hypothesis that the disease is present. The probability of the infection after the physician reviews evidence for a UTI is

$$q(\text{UTI} \mid \text{H\&P}) = ||P(h_1)\alpha_{\text{h\&p}}||^2 = .7929.$$

The revised state $\alpha_{\text{h\&p}}$ represents the physician's beliefs from the history and physical perspective. As we can see from above, the probabilities associated with states $h_1 \wedge e_2$ and $h_2 \wedge e_2$ are zero from this perspective. However, the physician believes that the evidence will look different when he or she changes perspective. Quantum probability theory allows for an amplitude to change from zero to a non-zero value. This cannot happen in a Bayesian model because once a state is certain, it remains certain. According to the Bayesian model, if we condition the probability of the infection on the presence of e_1, the physician will believe with certainty that e_1 is true. However, in the quantum model, the uncertainty principle states that certainty from one perspective implies uncertainty from another perspective. Thus, the physician wants to view the evidence from both perspectives.

When the physician reviews the laboratory data, he or she finds evidence for the absence of the UTI. As we assume that the history and physical event and the laboratory work event are incompatible, we must perform a change of basis by applying the $U_{BN}U_{AN}^{\dagger}$ unitary transformation to the state $\alpha_{\text{h\&p}}$:

$$\beta_{\text{h\&p}} = U_{BN}U_{AN}^{\dagger}\alpha_{\text{h\&p}} \Rightarrow \begin{bmatrix} .6054 \\ .1159 \\ .1645 \\ .1141 \end{bmatrix}.$$

The laboratory data provide no evidence of the UTI, so we project the state $\beta_{\text{h\&p}}$ onto the negative evidence for the disease or e_2 subspace:

$$P(e_2)\beta_{\text{h\&p}} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \beta_{\text{h\&p}} \Rightarrow \begin{bmatrix} 0 \\ .1159 \\ 0 \\ .1141 \end{bmatrix}.$$

The state after the presentation of the laboratory work is

$$\beta_{\text{h\&p,lab}} = \frac{P(e_2)\beta_{\text{h\&p}}}{||P(e_2)\beta_{\text{h\&p}}||} \Rightarrow \begin{bmatrix} 0 \\ .5039 \\ 0 \\ .4961 \end{bmatrix}.$$

The probability of infection from this final point of view is

$$q(\text{UTI} \mid \text{H\&P, Lab}) = ||P(h_1)\beta_{\text{h\&p,lab}}||^2 = .5039.$$

In general, the quantum model accounts for order effects by using different sequences of unitary transformations for different orderings of information. To calculate $q(h_1 \mid A, B)$, we first ''rotate'' the initial basis to the A basis by U_{AN}. Then, we ''rotate'' the A basis to the B basis by U_{BA}. However, if we were to calculate $q(h_1 \mid B, A)$, we would transform the initial basis to the B basis by U_{BN}. Next, we would transform the B basis to the A basis by U_{AB}. (See Fig. 4.) Even though we start with the initial basis for both cases, the quantum model ends in the B basis for one calculation and the A basis for the other. The state vectors, β_{e_1,e_2} and α_{e_2,e_1}, are used in the final calculations, $q(h_1 \mid A,B)$ and $q(h_1 \mid B,A)$, respectively. Order

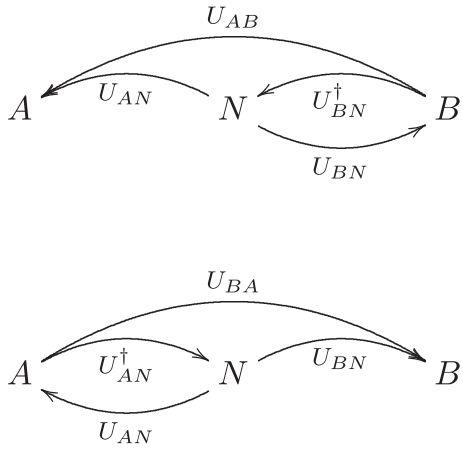

Fig. 4. Sequence of unitary transformations for $q(h_1 \mid B,A)$ (top) and $q(h_1 \mid A,B)$ (bottom).

effects arise from the quantum model because the projectors do not commute with the unitary matrices.

4. Jury decision making

Now, we turn to a set of experiments involving both manipulations of order and evidence strength in jury decision-making tasks. McKenzie et al. (2002) examined changes in confidence of guilt on hearing two sides of a criminal case. Specifically, participants were asked to imagine that they were jurors involved in a criminal trial and to rate their confidence in guilt (G) after reading the prosecution's case and the defense's case. The order and strength of the cases were manipulated between groups. A separate group rated the strength of the prosecution and defense on a 21-point scale with +10 indicating strong evidence of guilt and −10 indicating strong evidence of innocence.

In the first experiment, the participants were told that a warehouse was burglarized in the middle of the night. The next day, the defendant was arrested after the police received an anonymous tip. One group of subjects was presented with a strong prosecution (SP) followed by a weak defense (WD) and another group was presented with a weak defense followed by a strong prosecution. The summary of the prosecution's case was as follows:

- A witness identified the defendant in a police lineup.
- The type of burglary (commercial) is the same as that of the defendant's past convictions.
- The method of entry into the warehouse was unusual and is identical to that of the defendant's past convictions.

The summary of the defense's case read:

- A friend of the defendant's testified that he was with the defendant at his house at the time of the burglary.

For this experiment, the independent ratings of the cases showed that the prosecution was viewed as fairly strong ($M = 5.8$) and the defense was viewed as very weak ($M = -0.06$). The subjects rated their confidence in guilt on a 21-point scale with 0 = ''certain of innocence,'' 10 = ''guilt as likely as innocence,'' and 20 = ''certain of guilt.''

In a second experiment, McKenzie et al. varied the strength of the prosecution between two groups. One group rated their confidence in guilt after reading the strong prosecution followed by the weak defense, whereas another group rated their confidence in guilt after reading a weak prosecution (WP) followed by the same weak defense. The weak prosecution read:

- The method of entry into the warehouse was unusual and is identical to that of the defendant's past convictions.

In this experiment, the weak defense increased confidence in guilt when preceded by a strong prosecution but decreased confidence in guilt when preceded by a weak prosecution.

Specifically, $\Pr(G \mid \mathrm{SP}) < \Pr(G \mid \mathrm{SP}, \mathrm{WD})$ and $\Pr(G \mid \mathrm{WP}) > \Pr(G \mid \mathrm{WP}, \mathrm{WD})$. These results imply that the same weak defense provides positive evidence for guilt in one group and negative evidence for guilt in the other. By averaging the data from the prosecution-then-defense group from Experiment 1 with the strong-prosecution group from Experiment 2 and converting the confidence ratings to probabilities, we have the results shown in Table 3. In both experiments, the mean probability judgment of guilt before the presentation of either case was approximately .5.

4.1. Fitting the quantum model

We now demonstrate the quantum inference model's ability to account for order effects related to changes in information strength. To start, we let $h_1 =$ guilty and $h_2 =$ not guilty. Information A corresponds to the prosecutor's cases and information B corresponds to the defense's case. Note that in some conditions, information A corresponds to the strong prosecution (SP). However, in other conditions, information A corresponds to the weak prosecution (WP). Regardless, A provides positive evidence e_1 for guilt h_1. On the other hand, the defense B provides evidence e_2 for innocence h_2.

Like the medical example before, we must specify the quantum model's priors before we can proceed with the inference calculations. We assume that jurors have no initial preference toward guilt or innocence. Thus, our initial state vector is given by:

$$\omega = \begin{bmatrix} \omega_{h_1,e_2} \\ \omega_{h_1,e_2} \\ \omega_{h_2,e_1} \\ \omega_{h_2,e_2} \end{bmatrix} = \frac{1}{\sqrt{4}} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

We fit the quantum inference model to the average data from McKenzie's first and second experiments given in Table 3. We use three parameters to fit the six conditional probabilities given in the table. Specifically, x_{SP} coordinates the juror's belief in the defendant's guilt with the strong prosecution, x_{WP} coordinates the juror's belief in guilt with the weak prosecution, and x_{WD} coordinates the juror's belief in innocence with the weak defense. Although three parameters to six data points is not a high ratio of data points to parameters, the data are very complex and remain a challenge to fit even with many parameters. The results are shown in Table 4. The best-fitting parameters are $x_{\mathrm{SP}} = 2.3238$, $x_{\mathrm{WP}} = -3.5370$, $x_{\mathrm{WD}} = -1.8154$.

Table 3
Mean probability of guilt from Experiments 1 and 2

After First Case	After Second Case
$\Pr(\mathrm{G} \mid \mathrm{SP}) = .672$	$\Pr(\mathrm{G} \mid \mathrm{SP}, \mathrm{WD}) = .719$
$\Pr(\mathrm{G} \mid \mathrm{WD}) = .51$	$\Pr(\mathrm{G} \mid \mathrm{WD}, \mathrm{SP}) = .75$
$\Pr(\mathrm{G} \mid \mathrm{WP}) = .600$	$\Pr(\mathrm{G} \mid \mathrm{WP}, \mathrm{WD}) = .525$

Table 4
Quantum inference model results for jury task

After First Case	After Second Case
Pr(G | SP) = .688	Pr(G | SP, WD) = .703
Pr(G | WD) = .506	Pr(G | WD, SP) = .745
Pr(G | WP) = .598	Pr(G | WP, WD) = .509
SSE	0.00056

4.2. The belief-adjustment model

To further evaluate the quantum inference model, we compared the quantum model with an alternative explanation of order effects, the belief-adjustment model (Hogarth & Einhorn, 1992). The belief-adjustment model accounts for order effects by either adding or averaging evidence.

The belief-adjustment model assumes that individuals update beliefs by a sequence of anchoring-and-adjustment processes. The algebraic description of the model is

$$C_k = C_{k-1} + w_k \cdot (s(x_k) - R), \tag{2}$$

where $0 \le C_k \le 1$ is the degree of belief in the defendant's guilt after reading case k, $s(x_k)$ is the strength of case k, R is a reference point, and $0 \le w_k \le 1$ is an adjustment weight for case k. Hogarth and Einhorn (1992) argued that evidence can be encoded either in an absolute manner or in relationship to the current belief in the hypothesis. If evidence is encoded in an absolute manner and there exists a positive/negative relationship between the evidence and hypothesis, $R = 0$ and $-1 \le s(x_k) \le 1$. However, if evidence is encoded in relationship to the current belief, $R = C_{k-1}$ and $0 \le s(x_k) \le 1$. Also, Hogarth and Einhorn assumed that the adjustment weight w_k depends on the level of current belief and the sign of the difference $s(x_k)-R$. Specifically, if $s(x_k) \le R$, then $w_k = C_{k-1}$. However, if $s(x_k) > R$, then $w_k = 1-C_{k-1}$.

Using this information, we can rewrite the belief-adjustment model as either an adding model or an averaging model. The adding model results when information is encoded in an absolute manner and is given by

$$C_k = \begin{cases} C_{k-1} + C_{k-1} \cdot s(x_k), & \text{if } s(x_k) \le 0 \\ C_{k-1} + (1 - C_{k-1}) \cdot s(x_k), & \text{if } s(x_k) > 0 \end{cases}.$$

In contrast, the averaging model results when information is encoded in relationship to the current belief and is given by

$$C_k = \begin{cases} C_{k-1} + C_{k-1} \cdot (s(x_k) - C_{k-1}), & \text{if } s(x_k) \le C_{k-1} \\ C_{k-1} + (1 - C_{k-1}) \cdot (s(x_k) - C_{k-1}), & \text{if } s(x_k) > C_{k-1} \end{cases}.$$

Rearranging the terms above shows that the current belief is an average of the prior belief and the strength of the new evidence weighted by the prior belief.

McKenzie et al. (2002) showed that the standard belief-adjustment model cannot account for the probabilities given in Table 3. This is due to the fact that the weak defense increased confidence in guilt when preceded by the strong prosecution but decreased confidence in guilt when preceded by the weak prosecution. McKenzie et al. developed an extended version of the belief-adjustment model called the minimum acceptable strength (MAS) model to account for the results.

The MAS model extends the belief-adjustment model by defining the reference point as a case's MAS (McKenzie et al., 2002). Thus, Eq. (2) becomes

$$C_k = C_{k-1} + w_k \cdot (s(x_k) - m_{k-1}), \tag{3}$$

where m_{k-1} is the MAS of the previous case and $-1 \leq s(x_k) \leq 1$. Neither the adding model nor the averaging model can predict that a defense would increase confidence in guilt. However, it is possible to select a value for m_{k-1} such that the difference between the strength of the weak defense and m_{k-1} is positive. Therefore, confidence in guilt increases as a result of the weak defense. McKenzie et al. argued that a strong case presented first produces a demanding reference point, leaving room for a later weak case to fall short. For example, if the strength of a weak defense is −0.1 and the MAS determined by the previous case is −0.3, then the weak defense will fall short and have a reverse impact on judgments of guilt. The downside to the MAS model is the increase in parameters. The adding and averaging models specify a parameter for each case, namely $s(x_k)$. However, the MAS model also needs a MAS parameter for each case, thereby doubling the number of parameters needed in the original model.

We fit McKenzie's model of MAS to the data in Table 3. We assumed $-1 \leq s(x_k)$, $m_j \leq 1$, and we let the prior confidence in guilt be $C_0 = 0.5$. Following Hogarth and Einhorn (1992), we let

$$w_k = \begin{cases} C_{k-1}, & \text{if } s(x_k) \leq m_{1-k} \\ 1 - C_{k-1}, & \text{if } s(x_k) > m_{1-k} \end{cases}.$$

We set the MAS of the first case to be $m_0 = 0$ because we assume the juror has no prior assumptions about case strength. Because $s(x_k) \in [-1,1]$ for all k, we used a logistic function to map the average independent strength ratings from participants into this desired range. Specifically, the strength, $s(x_k)$, of case k is determined by

$$s(x_k) = -1 + \frac{2}{1 + e^{-(Bx_k)}},$$

where B is a gradient parameter and x_k is the average independent strength rating for the case in question. Besides the gradient parameter, McKenzie's model requires three MAS parameters used in the calculation of Pr(G | SP, WD), Pr(G | WD, SP), and Pr(G | WP, WD). We denote these three MAS parameters as m_{SP}, m_{WD}, and m_{WP}, respectively. The results are shown in Table 5. The best-fitting parameters are $B = 1.3161$, $m_{\mathrm{SP}} = -0.1562$, $m_{\mathrm{WD}} = -0.1548$, and $m_{\mathrm{WP}} = 0.0541$.

Table 5
Minimum acceptable strength model results for jury task

After First Case	After Second Case
Pr(G | SP) = .682	Pr(G | SP, WD) = .719
Pr(G | WD) = .480	Pr(G | WD, SP) = .750
Pr(G | WP) = .565	Pr(G | WP, WD) = .513
SSE	0.0022

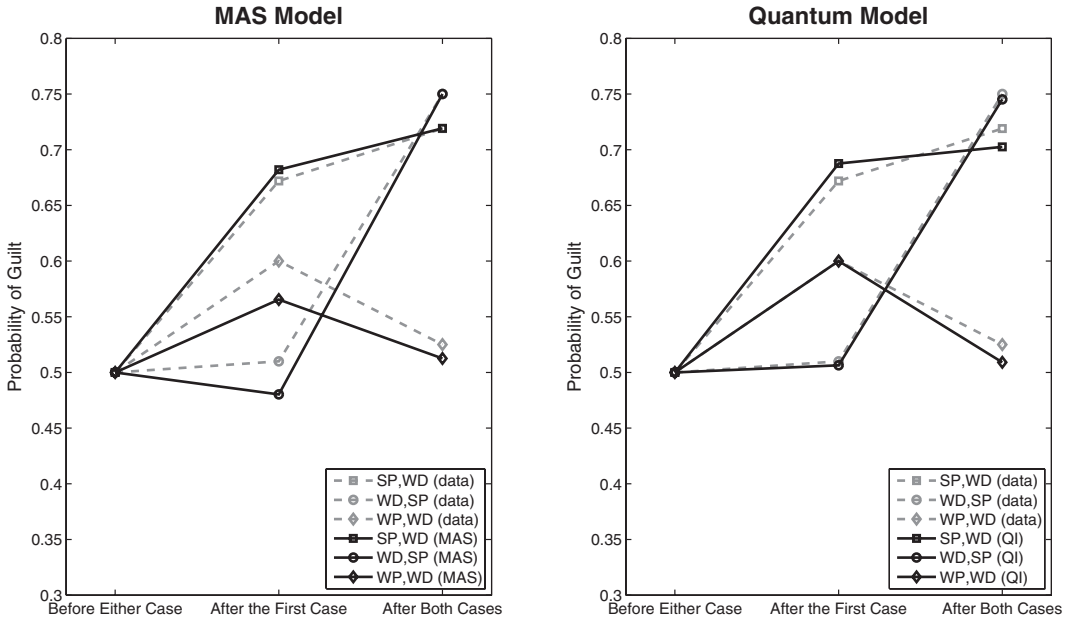

Fig. 5. MAS and quantum model fits to the mean probability of guilt from the study conducted by McKenzie et al. (2002).

Fig. 5 illustrates the MAS and quantum model fits to the probabilities given in Table 3. From the figure, we see that the quantum model provides a better fit to the data. Further, we rescaled the dependent variable to lie between 0 and 100 and calculated the root mean-squared error to provide an estimate of the average error on a 0–100 scale for each model. These calculations show that the average error for the MAS model is about 0.19, whereas the average error for the quantum model is only about 0.10.

Another drawback to the MAS model is the number of parameters needed to fit the data. We used four parameters to fit the MAS model, whereas we used three parameters to fit the quantum model. Moreover, this number increases to 7 if case strengths $s(x_k)$ are treated as free parameters.

Table 6
Conditional probabilities from inference task

After First Case	After Second Case	
Pr(G | WP)	Pr(G | WP, WD)	Pr(G | WP, SD)
Pr(G | SP)	Pr(G | SP, WD)	Pr(G | SP, SD)
Pr(G | WD)	Pr(G | WD, WP)	Pr(G | WD, SP)
Pr(G | SD)	Pr(G | SD, WP)	Pr(G | SD, SP)

5. Experiment 1: Extending McKenzie et al.

McKenzie et al. did not examine all possible combinations of case strength and order. Assuming that there are only two possible strengths, weak and strong, there are 12 total possible conditional probability judgments that can be made (see Table 6). Thus, we designed a new experiment to collect data for these 12 probabilities. Participants in this new study read eight different scenarios involving a defendant standing trial for either robbery, larceny, or burglary. Each participant was placed into one of eight conditions for each scenario. These eight conditions arise from the eight possible sequential judgments that can be made when taking into consideration order and case strength (e.g., weak prosecution followed by strong defense). Participants were placed in a different condition for each crime so they would experience all eight conditions by the end of the experiment. The participants reported the likelihood of the defendant's guilt before reading either case, after the first case, and after the second case.

Participants in the study were 299 undergraduate students from Indiana University who received experimental credit for introductory psychology courses. For each scenario, there were approximately 38 participants in each condition. All stimuli were presented on a computer and students entered their responses using the computer keyboard. For each scenario, participants were asked to imagine that they were jurors on the trial. They were also told that in each crime, the defendant was arrested after the police received an anonymous tip. One of the eight scenarios was directly patterned after the crime used by McKenzie et al. Likelihood of the defendant's guilt was reported on a continuous scale from 0 to 1 with 0 = ''certain not guilty,'' 0.5 = ''equally likely'', and 1 = ''certain guilty.''

5.1. Results

Eight of the 299 participants were excluded from the analyses because the majority of their initial ratings (before being presented with the prosecution or defense) were 0. These participants most likely assumed a literal interpretation of ''innocent until proven guilty.''

We first analyzed each scenario alone, and our analysis revealed a prevalence of recency effects. These effects arise when decision makers place disproportionate importance on recent evidence (e.g., $\Pr(G \mid SP, SD) < \Pr(G \mid SD, SP)$). For each crime, there were four defense-prosecution pairs (SD vs. SP, SD vs. WP, WD vs. SP, and WD vs. WP) that could exhibit order effects. A two-sample t test showed the majority of pairs exhibited a significant recency effect ($p < .05$). Because the scenarios were designed to be very similar, we

reanalyzed the data by collapsing across all eight scenarios. A two-sample t test showed a significant recency effect for each of the four defense-prosecution pairs ($p < .001$).[4]

5.2. Fitting the data

The presence of recency effects in this new data set confirms earlier findings and provides the largest data set so far for comparing models that explain recency effects. As mentioned, Hogarth and Einhorn discovered that recency effects are prevalent in simple, step-by-step tasks with short series of evidence. Furthermore, there is evidence of recency effects in studies involving mock trials (Furnham, 1986; Walker et al., 1972).

As the data exhibit recency effects, we can fit the standard belief-adjustment model instead of the MAS model. We fit the averaging model, the adding model, and the quantum inference model to the mean likelihood of guilt for the eight different crimes as well as the averaged data. All three models used four parameters to fit the 12 data points associated with each crime. These parameters were fit by minimizing the SSE between the data and model predictions. The four parameters used by the averaging and adding models arise from the four case strengths, $s(x_k)$, in Eq. (2). The four parameters for the quantum model arise

Table 7
Estimated average errors for three models

Crime	Averaging	Adding	Quantum
1	0.77	**0.31**	0.33
2	0.73	0.26	**0.22**
3	1.00	0.42	**0.24**
4	0.88	0.36	**0.33**
5	0.87	**0.28**	0.30
6	0.74	0.33	**0.31**
7	0.87	0.42	**0.27**
8	0.72	0.37	**0.14**
Average	0.77	0.22	0.22

Table 8
Model fits for eight order conditions

Data		Averaging		Adding		Quantum	
First	Second	First	Second	First	Second	First	Second
WP = 0.651	WP,WD = 0.516	0.578	0.552	0.631	0.541	0.647	0.502
	WP,SD = 0.398		0.436		0.404		0.407
SP = 0.805	SP,WD = 0.687	0.748	0.587	0.793	0.680	0.870	0.689
	SP,SD = 0.540		0.437		0.508		0.527
WD = 0.390	WD,WP = 0.619	0.499	0.589	0.394	0.587	0.390	0.639
	WD,SP = 0.779		0.747		0.768		0.758
SD = 0.278	SD,WP = 0.495	0.401	0.568	0.294	0.519	0.275	0.487
	SD,SP = 0.690		0.756		0.730		0.702

from the matrix operators used to transform the belief vector. The root mean-squared errors rescaled between 0 and 100 for all three models are shown in Table 7. From this table, we see that there is less error for both the adding and quantum models than the averaging model. Further, the quantum model fits slightly better than the adding model in most cases.

Table 8 shows the mean judgment, averaged over participants for the eight cases. The initial judgment produced a mean probability equal to .459. The first column of Table 8 shows the effect of the first piece of information (i.e. , either the prosecution or defense). From the table, we see that there is a clear effect of evidence strength on probability judgments. The second column of the table shows probability judgments after both pieces of information are presented. From the data in this column, we see strong recency effects. The remaining columns provide the estimates from the averaging model, the adding model, and the quantum model.

5.3. A generalization test

To further assess the three models (quantum, adding, and averaging), we collected data from 432 new subjects completing the same jury decision-making task as discussed above. In this version of the experiment, subjects gave responses on the 21-point confidence scale used by McKenzie et al. (2002). Other than the change in reporting scale, there were no other differences in the two versions of the task. Similar to the first version of the task, we analyzed the data by collapsing across all eight scenarios, and a two-sample t test showed a significant recency effect for each of the four defense-prosecution pairs ($p < .001$).[5] We fit the three models by minimizing the SSE. The rescaled root mean-squared error was 0.27 for the quantum model, 0.31 for the adding model, and 0.89 for the averaging model.

We used the data collected from the two versions of the task to perform split-half cross-validation (Browne, 2000; Shiffrin, Lee, Wagenmakers, & Kim, 2008).[6] First, we used the data from the probability scale version of the task to form the training set and the data from the confidence rating version to form the test set. Next, we let the data from the confidence rating version form the training set and the data from the probability scale version form the test set. This procedure allowed us to assess model generalization across different response scales and populations (Busemeyer & Wang, 2000). We calculated the mean-squared error for each test set for all three models. The overall mean-squared error for each model was calculated by averaging the mean-squared error from the two test sets. The overall rescaled root mean-squared error was 0.32 for the quantum model, 0.36 for the adding model, and 0.85 for the averaging model. From these results, we conclude that the quantum model best predicts the unseen data followed closely by the adding model.

6. Experiment 2: Extreme evidence

To provide even more of a distinction between the quantum model and the adding model, we conducted a second jury decision-making experiment involving extreme evidence.

In this task, subjects read about an individual on trial for a crime in which the defense had an irrefutable argument. Specifically, the defense stated that the defendant was giving a public lecture when the crime was committed. The prosecution's argument was moderately strong: A witness claimed to have seen the defendant near the scene of the crime. It seems reasonable to believe that the probability of guilt after hearing the defense will be near zero. Now, if the prosecution is presented after the defense, it is unlikely that the probability of guilt will increase by much.

6.1. Model predictions

So far, the model comparisons indicate that the averaging model does not fit as well as the quantum model or the adding model, but the fits of the latter two are so similar that it is difficult to distinguish them. Therefore, we designed an experiment that makes a priori predictions to distinguish the two models. Because the predictions are a priori and parameter free, one cannot argue that a better prediction is merely a consequence of one model being more flexible than another. We designed a situation in which the adding model predicts that a piece of evidence (i.e., the prosecution) will produce a minimal effect when observed by itself, whereas the quantum model predicts that the same piece of evidence will have a major effect.

The problem for the adding model is that an extreme defense constrains the prosecutor's evidence to remain a very low strength even when it comes first; whereas for the quantum model, under the same constraint of extreme defense evidence, the prosecutor's evidence must remain strong when it comes first. To see this, let's first examine the adding model:

$$C_p = C_d + (1 - C_d) \cdot s(x_p),$$

where C_p is the evaluation of the guilty hypothesis after hearing the prosecution's case. We might assume the evaluation of the hypothesis after hearing just the defense, C_d, is near 0, say $C_d = \epsilon_1$. Thus, $s(x_p)$ must also be a near 0, say $s(x_p) = \epsilon_2$, in order for C_p to remain small:

$$C_p = \epsilon_1 + (1 - \epsilon_1) \cdot \epsilon_2 = \epsilon_1 + \epsilon_2 - \epsilon_1 \cdot \epsilon_2 \approx 0.$$

Now, suppose the prosecution is presented before the defense. According to the adding model,

$$C_p = C_0 + (1 - C_0) \cdot s(x_p),$$

where C_0 is the evaluation of the guilty hypothesis before hearing either the prosecution or defense. We might assume that $C_0 \approx 0.5$. Thus, we have

$$C_p = 0.5 + 0.5 \cdot \epsilon_2 \approx 0.5,$$

showing the prosecution has little impact on the initial evaluation of the hypothesis. However, it seems unlikely that initial beliefs will be unaltered by the presentation of

the prosecution. On the contrary, we might expect this prosecution to be very effective when no prior defense is presented. Essentially, the problem arises from the model's assumption that the strength of the prosecution, $s(x_p)$, is determined independently of other evidence.

We also derived a priori predictions from the quantum model by examining all the possible values for Pr(Guilty | Prosecution) that the model can produce. Because closed form mathematical predictions are difficult to derive, we gathered information about this probability by exploring the parameter space. Because we believe that the probability of guilt after hearing the defense will be near zero, we restricted our search to parameters yielding Pr(Guilty | Defense), Pr(Guilty | Defense, Prosecution) ≤ .1. We might also assume that the probability of guilt after hearing the prosecution followed by the defense will be small. So we also restricted our search to parameter values resulting in Pr(Guilty | Prosecution, Defense) ≤ .2.

The quantum model uses two parameters to account for all four probabilities mentioned before. Of over nine million parameter pairs, 1,340 pairs met all of the restrictions. From these parameters, we found that Pr(Guilty | Prosecution) can range from .5017 to .9965. Fig. 6 illustrates the results of the parameter search as a histogram showing the frequency of particular values of Pr(Guilty | Prosecution). From this graph, we see that probabilities near .5 rarely occur. In fact, the most frequent probabilities occur between .7 and .9. These results suggest the quantum model is a priori more likely to predict an increase in probability with the presentation of the prosecution.

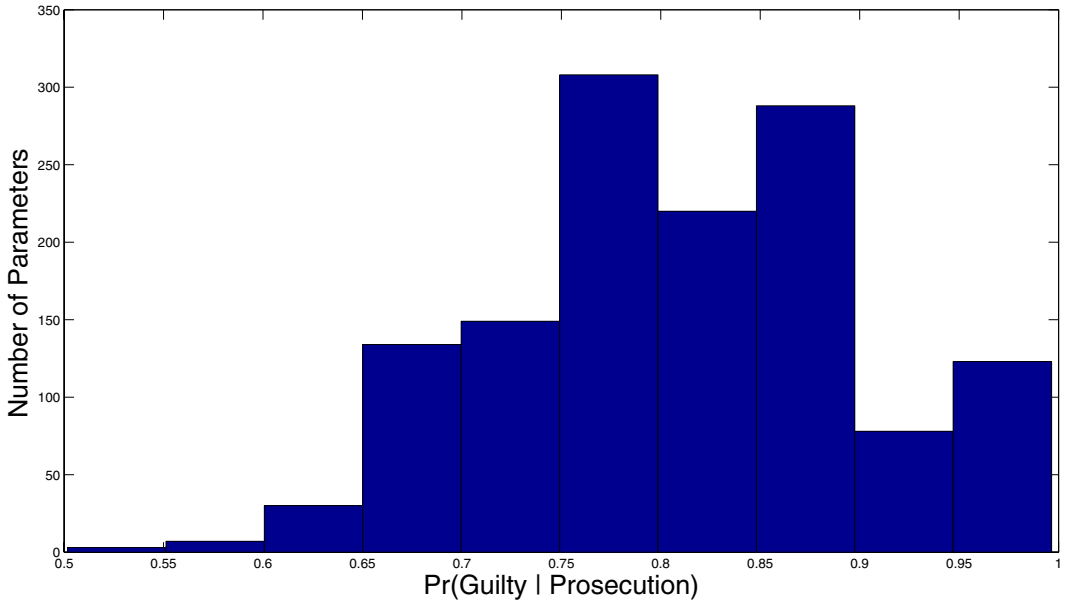

Fig. 6. Frequency of Pr(Guilty | Prosecution) from the quantum inference model.

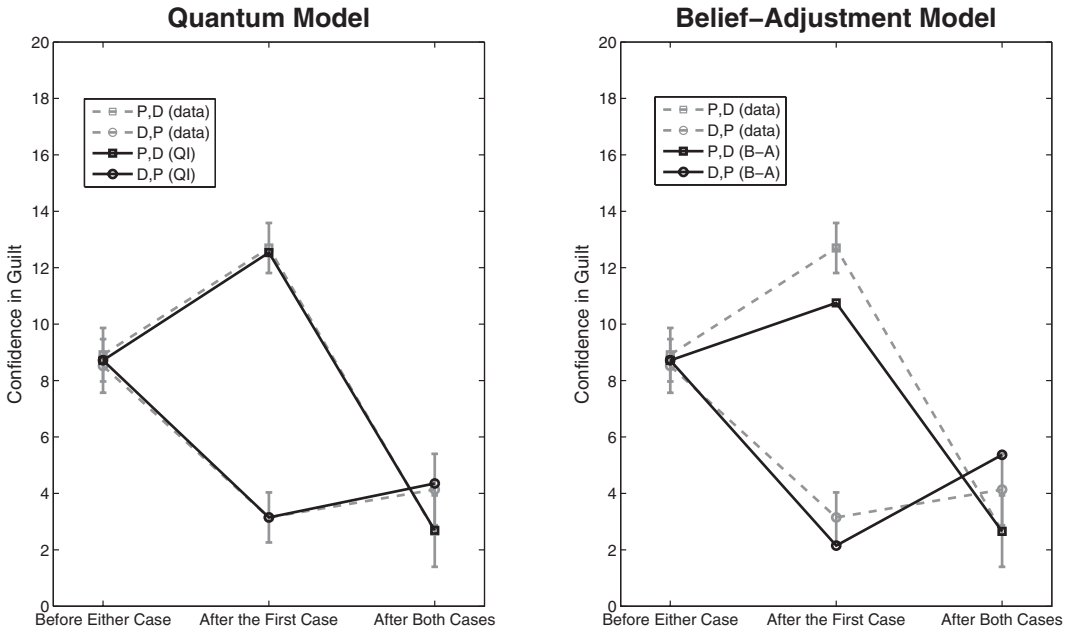

Fig. 7. Adding and quantum model fits to the mean probability of guilt from Experiment 2. Error bars on the data show the 95% confidence interval.

6.2. The experiment

Participants in the study were 164 undergraduate psychology students from Indiana University who received experimental credit for introductory psychology courses. Subjects were placed into one of two conditions corresponding to the two possible case orders: prosecution followed by defense or defense followed by prosecution. Similar to Experiment 1, subjects entered responses on a computer and were told that the defendant was arrested after the police received an anonymous tip. Instead of providing the likelihood of the defendant's guilt, subjects were asked to rate their confidence in guilt on the same 21-point scale used by McKenzie et al. Like Experiment 1, a significant recency effect was found ($p < .023$).[7]

We converted the confidence ratings to probabilities and fit the quantum model and the adding model to the mean of these probabilities. We did not fit the averaging model because Experiment 1 shows the adding model outperforms the averaging model. Fig. 7 shows the model fits for the two models. Both the quantum model and the adding model use two parameters (x_P and x_D for the quantum model and $s(x_p)$ and $s(x_d)$ for the adding model) to fit the data. We calculated the rescaled root mean-squared error to provide an estimate of the average error for each model. These calculations show that the average error for the adding model is about 0.62, whereas the average error for the quantum model is only about 0.07.

The standard belief-adjustment model cannot capture dependences between the strength of the prosecution and the irrefutable defense. As predicted, it substantially underestimates

the impact of the positive evidence when it comes first. Thus, the model provides a poor fit to the data. Unlike the belief-adjustment model, the quantum model does not assume individuals combine evidence by simple arithmetic procedures such as adding or averaging. The quantum model correctly predicts a large impact of the prosecution when it comes first. The quantum model accounts for the effects of extreme evidence by assuming that judges view information from different or incompatible perspectives.

7. General discussion

Cognitive models based on the principles of quantum probability have the potential to explain paradoxical phenomena arising in cognitive science. Previously, quantum models have been used to account for violations of rational decision-making principles (Pothos & Busemeyer, 2009), paradoxes of conceptual combination (Aerts, 2009), human judgments (Khrennikov, 2004), and perception (Atmanspacher, Filk, & Romer, 2004). This article describes a quantum inference model that accounts for order effects observed in human inference tasks involving two hypotheses.

The quantum model models order effects in terms of the change in viewpoints produced by evaluating different sources of information. For example, in the jury decision task, the individual evaluates his or her belief state from different points of view depending on whether the source is the prosecutor of the defense. During the prosecutor's presentation, the person views the evidence from the prosecutor's point of view, and the belief state is evaluated with respect to the prosecution basis vectors (prosecution arguments); but later during the defense's presentation, the person changes to the defense's point of view and evaluates his or her beliefs with respect to the defense basis vectors (defense arguments). The rotations used to change from one perspective to another do not commute with the projectors, so that the order of performing these changes in perspectives matters, which then produces the order effects.

One might question the extent to which quantum probabilities are rational. Like classic (Kolmogorov/Bayesian) probability theory, quantum theory is based on a coherent set of axioms. Then, the question falls back on which set of axioms is most appropriate for an application. A Bayesian model assumes that all events are compatible. In other words, evaluating two events always results in the same outcome regardless of the order of evaluation. While this might be an appropriate assumption in many contexts, it is not clear why this should always be the case. Quantum probability theory gives modelers more freedom when approaching problems in which order manners. In quantum theory, events can be defined as either compatible or incompatible. In the case when all events are compatible, quantum probability is identical to classical probability. Deciding when two events should be treated as compatible or incompatible is an important research question. There has been some work on this problem for questions involving human judgments (Busemeyer, Pothos, Franco, & Trueblood, 2011).

Another important consideration is whether a set of axioms is within the range of human information processing capabilities. Classic probability theory requires assigning

probabilities to all combinations of hypotheses, types of evidence, and levels of evidence. For example, if there are two hypotheses, n sources, and two types of evidence per source, then the classic model requires assigning 2^{n+1} joint probabilities; but the quantum model presented here can maintain a low four-dimensional space by evaluating the n different sources from n different points of view. The classic model can attempt to overcome this curse of dimensionality by imposing additional assumptions (e.g., conditional independence), but this could sacrifice its normative status if the assumptions are invalid. Thus, the quantum approach may be the most coherent way to assign probabilities that remain within human information processing capabilities.

To summarize, we demonstrated that the quantum model can account for recency effects and order effects caused by the manipulation of information strength. These order effects violate the commutative axiom of Bayesian models. We first showed the quantum inference model can closely fit data from the medical decision-making task by Bergus et al. (1998). The quantum model accounts for the presence of recency effects in the medical task by using different viewpoints (sequences of basis vectors) to represent different orderings of information. Using data collected by McKenzie et al., we showed that the quantum inference model outperforms the MAS model. We also provided evidence that the quantum model performs as well or slightly better than the belief-adjustment model when fitting data from Experiment 1. Finally, we described some of the limitations of the belief-adjustment model in relationship to irrefutable evidence. We argued that the quantum inference model is not faced with these limitations and provides more reasonable predictions. In future work, we plan to study the complexity of the quantum model in a more rigorous fashion than simply equating the number of parameters in each model. For the purposes of this article, we felt that comparing the number of data points and parameters was a useful start or beginning point until a more detailed analysis can be performed. In the future, we plan to continue empirically investigating the quantum inference model in the hope of developing a more coherent theory concerning human inference tasks.

Notes

1. There is another line of research that uses quantum physical models of the brain to understand consciousness (Hammeroff, 1998) and human memory (Pribram, 1993). We are not following this line of work. Instead, we are using quantum models at a more abstract level analogous to Bayesian models of cognition.
2. We use Dirac, or Bra-ket, notation in keeping with the standard notation used in quantum mechanics. For our purposes, $|e_j\rangle$ corresponds to a column vector, whereas $\langle e_j|$ corresponds to a row vector.
3. The matrix exponential function is commonly available in mathematical programs, such as R, or Matlab, or Gauss, etc.
4. We used a two-sample t test without the assumption of equal variances. The t-statistic and degrees of freedom given by Satterthwaite's approximation for the four pairs are $t = -7.76$ and $df = 570.71$ for SD versus SP, $t = -4.89$ and $df = 579.99$ for SD

versus WP, $t = -5.82$ and $df = 562.55$ for WD versus SP, and $t = -5.93$ and $df = 567.76$ for WD versus WP.

5. Again, we used a two-sample t test without the assumption of equal variances. The t-statistic and degrees of freedom given by Satterthwaite's approximation for the four pairs are $t = -6.82$ and $df = 861.79$ for SD versus SP, $t = -4.47$ and $df = 859.55$ for SD versus WP, $t = -6.06$ and $df = 819.24$ for WD versus SP, and $t = -4.70$ and $df = 861.51$ for WD versus WP.
6. Although split-half cross-validation is not as good as using accumulation prediction error, the latter method is computationally difficult to implement with the quantum model because the parameter search is difficult.
7. We used a two-sample t test without the assumption of equal variances. The t-statistic and degrees of freedom given by Satterthwaite's approximation are $t = -2.30$ and $df = 152.11$.

Acknowledgments

This research was supported by the National Science Foundation/IGERT Training Program in the Dynamics of Brain-Body-Environment Systems at Indiana University and by the National Science Foundation under grant no. 0817965.

References

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

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

Anderson, N. H., & Hubert, S. (1963). Effects of concomitant verbal recall on order effects in personality impression formation. *Journal of Verbal Learning and Verbal Behavior*, *2*, 379–391.

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

Bergus, G. R., Chapman, G. B., Levy, B. T., Ely, J. W., & Oppliger, R. A. (1998). Clinical diagnosis and order of information. *Medical Decision Making*, *18*, 412–417.

Browne, M. W. (2000). Cross-validation methods. *Journal of Mathematical Psychology*, *44*, 108–132.

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

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

Franco, R. (2009). The conjunctive fallacy and interference effects. *Journal of Mathematical Psychology*, *53*, 415–422.

Furnham, A. (1986). The robustness of the recency effect: Studies using legal evidence. *Journal of General Psychology*, *113*, 351–357.

Gilovich, T., Griffin, D., & Kahneman, D. (2002). *Heuristics and biases: The psychology of intuitive judgment*. New York: Cambridge University Press.

Hammeroff, S. R. (1998). Quantum computation in brain microtubles? The penrose-hameroff ''orch or'' model of consciousness. *Philosophical Transactions Royal Society London (A)*, *356*, 1869–1896.

Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological Review*, *93*, 258–268.

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*, 1–55.

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

McKenzie, C. R. M., Lee, S. M., & Chen, K. K. (2002). When negative evidence increases confidence: Change in belief after hearing two sides of a dispute. *Journal of Behavioral Decision Making*, *15*, 1–18.

Nielsen, M. A., & Chuang, I. L. (2000). *Quantum computation and quantum information*. Cambridge, UK: Cambridge University Press.

Pothos, E. M., & Busemeyer, J. R. (2009). A quantum probability explanation for violations of ''rational'' decision theory. *Proceedings of the Royal Society B*, *276*(1165), 2171–2178.

Pribram, K. H. (1993). *Rethinking neural networks: Quantum fields and biological data*. Hillsdale, NJ: Erlbaum.

Shanteau, J. C. (1970). An additive model for sequential decision making. *Journal of Experimental Psychology*, *85*, 181–191.

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

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

Walker, L., Thibaut, J., & Andreoli, V. (1972). Order of presentation at trial. *Yale Law Journal*, *82*, 216–226.

Appendix: Eigen decomposition of unitary matrices

To provide some insight into the unitary matrices U_{AN} and U_{BN}, we proceed by computing the matrix exponential

$$U_{jN} = e^{-ix_j(H_1'' + H_2'')}, \tag{4}$$

where $j = A$ and B. First, we diagonalize the sum $H'' = H_1'' + H_2''$ such that $H'' = \Lambda \Gamma \Lambda^{-1}$, where

$$\Gamma = \begin{bmatrix} -1-\sqrt{3} & 0 & 0 & 0 \\ 0 & 1+\sqrt{3} & 0 & 0 \\ 0 & 0 & 1-\sqrt{3} & 0 \\ 0 & 0 & 0 & -1+\sqrt{3} \end{bmatrix},$$

is the matrix of eigenvalues and

$$\Lambda = \begin{bmatrix} \frac{1}{\sqrt{3}} & 2+\sqrt{3} & 2-\sqrt{3} & \frac{-1}{\sqrt{3}} \\ -1-\frac{2}{\sqrt{3}} & 1 & 1 & -1+\frac{2}{\sqrt{3}} \\ \frac{-1}{\sqrt{3}} & \sqrt{3} & -\sqrt{3} & \frac{1}{\sqrt{3}} \\ 1 & 1 & 1 & 1 \end{bmatrix},$$

is the matrix composed of eigenvectors. The eigenvectors are not normalized to maintain a simple numerical representation. Because U_{jN} is unitary, the Λ matrix is as well. As a result, the inverse of Λ is equal to its adjoint making it fairly easy to compute Λ^{-1}. Next, we can write Eq. (4) as

$$U_{jN} = \Lambda e^{-ix_j\Gamma}\Lambda^{-1},$$

where $e^{-ix_j\Gamma}$ is the matrix

$$\begin{bmatrix} \text{cis}[x_j(1+\sqrt{3})] & 0 & 0 & 0 \\ 0 & \text{cis}[x_j(-1-\sqrt{3})] & 0 & 0 \\ 0 & 0 & \text{cis}[x_j(-1+\sqrt{3})] & 0 \\ 0 & 0 & 0 & \text{cis}[x_j(1-\sqrt{3})] \end{bmatrix},$$

where $\text{cis}(x) = \cos(x) + i \sin(x)$.

This decomposition is useful in understanding the products of unitary matrices. For example, in the medical decision-making task, we assume that the group of physicians presented with the history and physical first change their point of view with the presentation of the laboratory results by applying the U_{BA} unitary transformation to the state $\alpha_{\text{h\&p}}$. By using the eigen decomposition, we can write this as

$$\begin{aligned} \beta_{\text{h\&p}} &= U_{BN}U_{AN}^{\dagger}\alpha_{\text{h\&p}} = [\Lambda e^{-0.3306i\Gamma}\Lambda^{-1}][\Lambda e^{-4.4045i\Gamma}\Lambda^{-1}]^{-1}\alpha_{\text{h\&p}} \\ &= \Lambda e^{(-0.3306+4.405)i\Gamma}\Lambda^{-1}\alpha_{\text{h\&p}} = \Lambda e^{4.0744i\Gamma}\Lambda^{-1}\alpha_{\text{h\&p}} \Rightarrow \begin{bmatrix} .6054 \\ .1159 \\ .1645 \\ .1141 \end{bmatrix}. \end{aligned}$$

This shows how the parameters $x_A = 4.4045$ and $x_B = 0.3306$ are combined in defining U_{BA}.