# A Quantum Probability Approach to Human Causal Reasoning

**Jennifer S. Trueblood (jstruebl@uci.edu)**
Department of Cognitive Sciences, University of California, Irvine
Irvine, CA 92697 USA

**Emmanuel M. Pothos (emmanuel.pothos.1@city.ac.uk)**
Department of Psychology, City University London
London, EC1V 0HB, UK

**Abstract**

When people make inferences about causal situations with vague and imperfect information, their judgments often deviate from the normative prescription of classical probability. As a result, it is difficult to apply popular models of causal reasoning such as $\Delta P$ and causal power, which provide good accounts of behavior in casual learning tasks and tasks where statistical information is provided directly. We propose a unified explanation of human causal reasoning using quantum probability theory that can account for causal reasoning across many different domains. In our approach, we postulate a hierarchy of mental representations, from fully quantum to fully classical, that could be adopted for different situations. We illustrate our approach with new experiments and model comparisons.

**Keywords:** Causal reasoning, quantum probability

## Introduction

Everyday we face situations where we must reason about causes and effects. Often, causal relationships are well established through practice - when I plug in and turn on an electric kettle, water boils. In other situations, we must reason about novel situations where there is vague and imperfect information - if I vote for the new mayoral candidate, will property taxes be lowered next year?

In general, causal judgment tasks can be divided into three types: experienced, statistical descriptions, and linguistic descriptions (Shanks, 1991). In experienced situations, participants witness firsthand the relationships between causes and effects. In situations with statistical descriptions, participants are given summary information about the frequency that specific causes or groups of causes produce effects. In situations with linguistic descriptions, participants are asked to make causal inferences from narratives, stories, texts, etc. These tasks often involve vague and imperfect information about causes and effects. It is this last class of problems that is our focus.

Models based on classical probability theory such as $\Delta P$ (Jenkins & Ward, 1965) and power PC theory[1] (Cheng, 1997) have provided good accounts of causal reasoning in both experienced situations and those described by statistical information. However, there is evidence that people's judgments about causal systems with linguistic descriptions often deviate from the normative prescription of classical probability (Sloman & Fernbach, 2011; Trueblood & Busemeyer, 2012). As a result, it is difficult to use models such as $\Delta P$ and causal power in this domain, thereby producing a disparate account of casual reasoning across tasks.

We propose a unified explanation of human causal reasoning using quantum probability theory.[2] In our approach, we postulate a hierarchy of mental representations that could be adopted for different situations. Classical probability models such as causal power represent one class of models in our hierarchy.

## Experiment 1

We begin by introducing the experimental paradigm that we use to illustrate our modeling approach. This paradigm is based on one developed by Rehder (2003) to study causal reasoning with novel categories. In our task, participants are given a linguistic description of a novel category, Lake Victoria Shrimp, and asked to judge the likelihood that certain features cause others. Specifically, participants are given information about two independent features that can influence a third feature. The language used to describe the features and their relationships is purposely vague as many real life situations do not involve precise information.

We were interested to see whether the order in which information is presented in our task affected final judgments. Order effects are well established in inference tasks (Hogarth & Einhorn, 1992) and are difficult to explain using simple classical probability models such as $\Delta P$ and causal power. However, quantum probability theory has been successful in accounting for these effects (Trueblood & Busemeyer, 2011).

### Methods

Participants learned about a novel biological category, Lake Victoria Shrimp, that had three binary features: ACh neurotransmitter (high or low amount), sleep cycle (accelerated or decelerated), and body weight (normal or high). Participants were given information about the typicality of feature values. For example, they were told that "Most Lake Victoria Shrimp have a high amount of ACh whereas some have a low amount of ACh". Participants were also given the causal relationships between features. These relationships were described as one feature causing another. Specifically, the ACh neurotransmitter and sleep cycle were described as affecting body weight. The strengths of causal relationships were described using the terms "often" and "sometimes". For example, participants

[1] We will refer to power PC theory as causal power throughout the remainder of the paper.

[2] We use the mathematical formalism of quantum theory without the associated physical meaning.

were told that "An accelerated sleep cycle often causes a high body weight".

Participants first studied the three features and the typicality of their values. After studying this information, participants took a multiple-choice test with six questions that tested them on this knowledge. Participants were required to answer each question correctly before moving on to the next one. Next, participants studied the two causal relationships and took another multiple-choice test with eight questions testing them on this new knowledge. As before, participants were required to answer each question correctly before moving on to the next one. Finally, participants were asked to take a few minutes to review the features and relationships one more time. After they finished reviewing this information, they completed a third multiple-choice test with 10 questions. In this final test, participants were only given one opportunity to answer each question. Their score on this test was used to gauge how well they learned the features and causal relationships.

After completing the learning stage, participants were asked to make sequences of two or three judgments about a particular shrimp as they learned new information about it. For example, they might first learn "A Lake Victoria shrimp is caught" and asked "How likely is it for the shrimp to have a high body weight?". Then, they learned information about different causes revealed during a series of lab tests. For example, participants might read "After lab testing, you learn that the shrimp has a high quantity of ACh neurotransmitter. Given this new information, how likely is it that this shrimp has a high body weight?". For the final judgment in the sequence, participants might read, "After further observation in the lab, you also learn that the shrimp has a decelerated sleep cycle. Given this new information, how likely is it that this shrimp has a high body weight?".

Participants were randomly assigned to one of two order conditions. In one condition, participants were always asked to make judgments about the ACh neurotransmitter before making judgments about the sleep cycle. In the other condition, this ordering was reversed. The between subjects design ensured subjects did not make judgments for reverse orderings as these judgments could be influenced by memory. In both conditions, participants were asked to make two types of judgments - predictive and diagnostic. Predictive judgments involved judging the likelihood of an effect given a cause (e.g., the likelihood of a shrimp having a high body weight given a low amount of ACh). Diagnostic judgments involved judging the likelihood of a cause given an effect (e.g., the likelihood of a low amount of ACh given a high body weight). The order in which the sequences were presented was randomized across participants. As shown in table 1, each participant completed four predictive (top) and four diagnostic (bottom) sequences. In the table, the predictive effect (i.e., body weight) is denoted by $e$, and the two causes, ACh neurotransmitter and sleep cycle, are denoted by $x$ and $y$ respectively. A feature value of 1 denotes a high amount of ACh, an accelerated sleep cycle, and high body weight for the respective features. Likewise, a feature value of 0 denotes a low amount of ACh, a decelerated sleep cycle, and a normal body weight.

Table 1: Sequences of judgments in Experiment 1.

| Condition 1 | | | Condition 2 | | |
|---|---|---|---|---|---|
| $p(e_1)$ | $p(e_1\|x_1)$ | $p(e_1\|x_1,y_1)$ | $p(e_1)$ | $p(e_1\|y_1)$ | $p(e_1\|y_1,x_1)$ |
| $p(e_1)$ | $p(e_1\|x_1)$ | $p(e_1\|x_1,y_0)$ | $p(e_1)$ | $p(e_1\|y_0)$ | $p(e_1\|y_0,x_1)$ |
| $p(e_1)$ | $p(e_1\|x_0)$ | $p(e_1\|x_0,y_1)$ | $p(e_1)$ | $p(e_1\|y_1)$ | $p(e_1\|y_1,x_0)$ |
| $p(e_1)$ | $p(e_1\|x_0)$ | $p(e_1\|x_0,y_0)$ | $p(e_1)$ | $p(e_1\|y_0)$ | $p(e_1\|y_0,x_0)$ |
| | $p(x_1\|e_1)$ | $p(x_1,y_1\|e_1)$ | | $p(y_1\|e_1)$ | $p(y_1,x_1\|e_1)$ |
| | $p(x_1\|e_1)$ | $p(x_1,y_0\|e_1)$ | | $p(y_0\|e_1)$ | $p(y_0,x_1\|e_1)$ |
| | $p(x_0\|e_1)$ | $p(x_0,y_1\|e_1)$ | | $p(y_1\|e_1)$ | $p(y_1,x_0\|e_1)$ |
| | $p(x_0\|e_1)$ | $p(x_0,y_0\|e_1)$ | | $p(y_0\|e_1)$ | $p(y_0,x_0\|e_1)$ |

Participants were asked to enter their likelihood judgments about specific features as numbers between 0 and 100 in a text box after reading each question. They were told that a judgment of 0 implied that they were certain the shrimp did not have the feature, a response of 50 implied that the shrimp was equally likely to have the feature or not, and a response of 100 implied that they were certain the shrimp did have the feature. On each trial, there was a picture of a scale from 0-100 reminding participants of this information.

122 University of California, Irvine undergraduates participated in the experiment online at a time of their choosing for course credit. There were 60 participants in condition 1 (i.e., ACh neurotransmitter first) and 62 participants in condition 2 (i.e., sleep cycle first).

## Results

The average score on the 10 question multiple choice test was 7.8 indicating most participants correctly learned the feature values and causal relationships during the learning stage. Each multiple choice question had three possible responses. If participants guessed for all of the questions, their expected score would be 3.3. We excluded participants with a score less than 7 out of 10 because it is crucial that participants understand the causal structure. This criterion excludes 38 participants leaving 41 in condition 1 and 43 in condition 2.

We examined order effects in the predictive judgment sequences by comparing final probability judgments in the sequences across conditions. Figure 1 compares the mean judgments from both conditions. In each plot, the left-most judgment is the prior probability of the effect, the next judgment is the probability of the effect given one of the causes (either $x$ or $y$), and the right-most judgment is the probability of the effect given both causes. If order effects exists, then the final, right-most judgments will be different. Two of the four comparisons produced order effects. The judgment $p(e_1|x_1,y_0)$ was significantly different (t(82) = -3.04, p = 0.003) from the judgment $p(e_1|y_0,x_1)$. Similarly, the judgment $p(e_1|x_0,y_1)$ was significantly different (t(82) = 5.51, $p < 0.001$) from the judgment $p(e_1|y_1,x_0)$. The type of order effect in these two comparisons is a recency effect which results from disproportionate importance of recent information. Judgments where

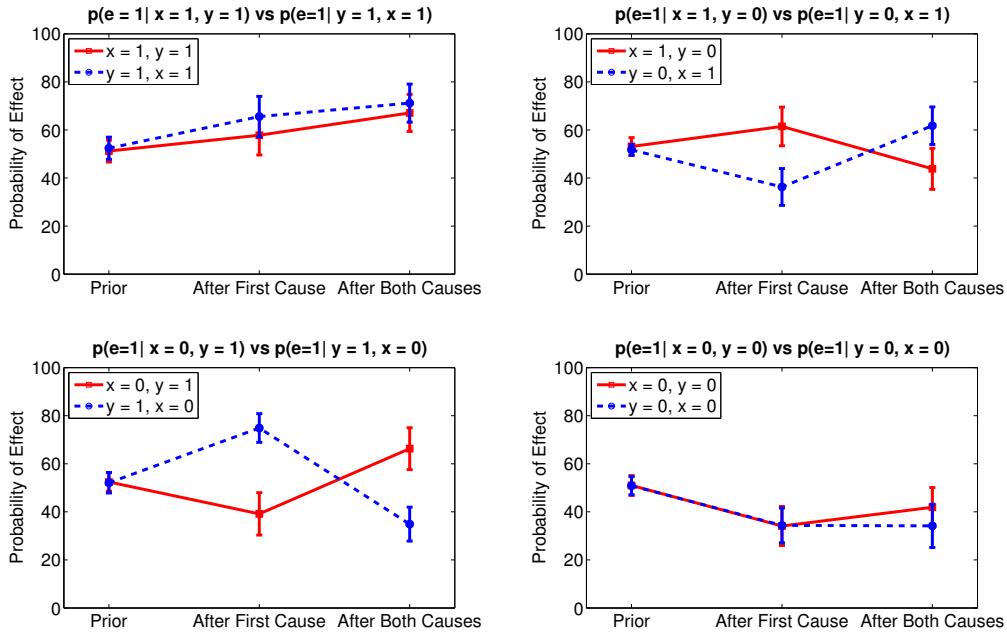


Figure 1: Mean judgments for the predictive sequences in Experiment 1. Condition 1 is shown in red and condition 2 in blue.

both feature values matched (i.e., both $x$ and $y$ equal 1 or both equal 0), did not produce significant order effects ($p > 0.05$).

Similar to the predictive sequences, we examined order effects in the diagnostic sequences by comparing final probability judgments in the sequences across conditions. Only the comparison of $p(x_0, y_0|e_1)$ to $p(y_0, x_0|e_1)$ produced a significant result ($t(82) = -2.68$, $p = 0.009$). The remaining three comparisons did not produce significant order effects ($p > 0.05$). This is consistent with findings showing violations of classical probability in predictive judgments, but not diagnostic judgments (Fernbach, Darlow, & Sloman, 2011).

It is possible that participants misinterpreted the second question so that they were judging $p(e_1|y)$ instead of $p(e_1|x, y)$ and similarly for the reverse ordering of $x$ and $y$. To control for this possible confound we ran a version of the experiment where all relevant feature information was included before each sequential judgment. For example, the second question in the predictive case might be "After further observation in the lab, you learn that the shrimp with a high quantity of ACh neurotransmitter also has a decelerated sleep cycle. Given this new information, how likely is it that this shrimp has a high body weight?". In experiment 1, participants were not reminded about the high quantity of ACh neurotransmitter before making the final judgment of the effect. As in experiment 1, predictive judgments where feature values mismatched produced significant recency effects. Thus, it is unlikely participants were misinterpreting the second question in experiment 1.

## A Hierarchy of Models

In classical probability theory, it is assumed there is a single space that provides a complete and exhaustive description of all events, which follows from the closure property of Boolean logic because if $x$ and $y$ are events in the sample space, so is the joint event $x \cap y$. Repeated application of this principle yields the elementary events of the sample space - those events that cannot be broken down any further. For example, consider a situation where there are three binary (1 or 0) events, $e$, $x$, and $y$ as in experiment 1. The elementary events of the sample space arise from the intersection of these three events and include events such as $e_1 \cap x_1 \cap y_0$. This sample space has $2^3 = 8$ elementary events. As the number of events increases, the size of the sample space rapidly increases. For only six binary events, the dimension of the sample space is 64. Plausibly, at some point, a person's capacity to combine events into a unified sample space and assign joint probabilities will be exceeded.

Quantum probability theory is a geometric approach to probability where events are represented as subspaces of a vector space. Unlike classical probability theory, quantum probability allows for multiple sample spaces, which are each associated with different bases for the same vector space. Different sample spaces (i.e., bases) are related geometrically by rotations. When two events are described by two different sample spaces, they are called *incompatible* and their joint event does not exist. Psychologically, this implies that individuals do not have a mental representation of the joint event. That is, they cannot think about these two events simultane-

ously. Rather, the events are processed in a sequence by first thinking of one event and then "rotating" to think about the other.

We hypothesize that individuals use multiple sample spaces when reasoning about unfamiliar causal problems with vague and imperfect information. Our hypothesis is motivated by the idea that causal learning is structurally local (Fernbach & Sloman, 2009). That is, when people are faced with a complex learning problem, they break the problem up by focusing on individual causal relationships rather than the full causal structure. Inferences about the full structure are constructed by combining local inferences piece by piece.

In experiment 1, participants learned about three binary features labeled $e$, $x$, and $y$ in table 1. There are at least three possible geometric approaches to modeling the features and their relationships using either two, four, or eight dimensions. These different approaches form a hierarchy of models associated with different types of mental representations.

In the 2-dimensional approach, we assume that individuals consider one feature at a time and do not have mental representations of joint events. Mathematically, each feature is represented as a different basis for a 2-dimensional vector space where the two dimensions correspond to the two possible feature values. This is equivalent to defining three separate samples spaces that each have two elementary events.

In the 4-dimensional approach, it is assumed that individuals form mental representations for single cause and effect relationships, but do not think about multiple causal relationships simultaneously. In this case, the two causal relationships are represented as different bases for a 4-dimensional vector space where the four dimensions are associated with the four joint events, $e_1 \cap x_1$, $e_1 \cap x_0$, $e_0 \cap x_1$, $e_0 \cap x_0$ for the causal relationship between $e$ and $x$ and similarly for $e$ and $y$. This is equivalent to defining two separate sample spaces that each have four elementary events.

In the 8-dimensional approach, it is assumed individuals have mental representations of all joint events. In this case, the eight dimensions are associated with the eight joint events such as $e_1 \cap x_1 \cap y_0$. This is equivalent to a classical probability model with a single sample space describing all events.

## 2-dimensional model

The 2-dimensional model assumes that the three binary features, $e$, $x$, and $y$, are each represented by different bases of a 2-dimensional vector space. Judgments of joint events such as $p(x \cap y)$ are formed by thinking of one feature first, say $x$, and then the other. That is, individuals do not think about the two features simultaneously.

Mathematically, the three bases associated with the three features are related to one another by rotation matrices. Because only the angle between bases matters, we can fix one of the bases and define the others in relationship to it. We set the basis vectors for $e$ to the standard basis for $\mathbf{R}^2$. Using Dirac notation, according to which $|v\rangle$ just denotes a column vector and $\langle v|$ is the conjugate transpose of this vector (i.e., the corresponding row vector), we have

$$|e_1\rangle = \begin{bmatrix}1\\0\end{bmatrix}, |e_0\rangle = \begin{bmatrix}0\\1\end{bmatrix}.$$

Then, we define the two rotation matrices

$$R_x = \begin{bmatrix}\cos(\theta_x) & -\sin(\theta_x)\\ \sin(\theta_x) & \cos(\theta_x)\end{bmatrix}, R_y = \begin{bmatrix}\cos(\theta_y) & -\sin(\theta_y)\\ \sin(\theta_y) & \cos(\theta_y)\end{bmatrix}$$

where the basis associated with feature $x$ is $\{R_x|e_1\rangle, R_x|e_0\rangle\}$ and the basis associated with feature $y$ is $\{R_y|e_1\rangle, R_y|e_0\rangle\}$.

Quantum probability theory postulates the existence of a state represented by a unit length vector $|\psi\rangle$ in the vector space. Psychologically, we assume this 2-dimensional state vector is a knowledge state that represents an individual's beliefs about the different features. Specifically, the two components of the state vector (i.e., probability amplitudes) correspond to the individual's beliefs about the two possible feature values, 1 and 0. Using probability amplitudes $\alpha_1$ and $\alpha_0$, we write the state vector as $|\psi\rangle = [\alpha_1, \alpha_0]'$ where $\alpha_0 = \sqrt{1-\alpha_1^2}$ in order to ensure the state vector is unit length. In total, the 2-dimensional model has three parameters: $\theta_x$, $\theta_y$, and $\alpha_1$.

Probabilities are calculated by projecting the state vector onto different subspaces. For example, the prior probability of the effect, $e_1$, is given by $p(e_1) = ||P_{e_1}|\psi\rangle||^2$ where $P_{e_1}$ is the 2 x 2 projection matrix with a 1 on the upper diagonal and zeros elsewhere. To calculate the predictive probability of the effect given a single cause such as $p(e_1|x_1)$, the state vector is first updated to a conditional state $|\psi_{x_1}\rangle = P_{x_1}|\psi\rangle / ||P_{x_1}|\psi\rangle||$ where the projector $P_{x_1}$ is defined by the outer product $|x_1\rangle\langle x_1|$ where $|x_1\rangle = R_x|e_1\rangle$. The probability of the effect is then calculated by projecting the conditional state onto the $e_1$ subspace: $||P_{e_1}|\psi_{x_1}\rangle||^2$. Calculations for other probabilities proceed in a similar manner.

## 4-dimensional model

The 4-dimensional model assumes that individuals have mental representations of single cause-effect relations so that joint events such as $e \cap x$ exist (i.e., the features can be thought about simultaneously). However, individuals do not have mental representations of joint events involving all three features such as $e_1 \cap x_1 \cap y_0$ .

In this model, two different 4-dimensional bases are used rather than three 2-dimensional bases. The two bases associated with the two causal relationships are related to one another by a rotation matrix. Because only the angle between bases matters, we can fix one of the bases and define the other in relationship to it. We set the basis vectors for the $x$-$e$ causal relationship to the standard basis for $\mathbf{R}^4$ and define the basis for the $y$-$e$ causal relationship with the rotation matrix $R = R_1 R_2 R_3$ where

$$R_1 = \begin{bmatrix}\cos(\theta_1) & -\sin(\theta_1) & 0 & 0\\ \sin(\theta_1) & \cos(\theta_1) & 0 & 0\\ 0 & 0 & \cos(\theta_1) & -\sin(\theta_1)\\ 0 & 0 & \sin(\theta_1) & \cos(\theta_1)\end{bmatrix}$$

$$R_2 = \begin{bmatrix} \cos(\theta_2) & 0 & -\sin(\theta_2) & 0 \\ 0 & \cos(\theta_2) & 0 & -\sin(\theta_2) \\ \sin(\theta_2) & 0 & \cos(\theta_2) & 0 \\ 0 & \sin(\theta_2) & 0 & \cos(\theta_2) \end{bmatrix}$$

$$R_3 = \begin{bmatrix} \cos(\theta_3) & 0 & 0 & -\sin(\theta_3) \\ 0 & \cos(\theta_3) & -\sin(\theta_3) & 0 \\ 0 & \sin(\theta_3) & \cos(\theta_3) & 0 \\ \sin(\theta_3) & 0 & 0 & \cos(\theta_3) \end{bmatrix}.$$

The matrix $R_1$ rotates the 2-dimensional subspaces associated with the different feature values of the effect. That is, the subspace associated with $e_1$ (i.e., events $e_1 \cap x_1$ and $e_1 \cap x_0$) and the subspace associated with $e_0$ (i.e., events $e_0 \cap x_1$ and $e_0 \cap x_0$). The matrix $R_2$ rotates the 2-dimensional subspaces associated with the two values of the cause. That is, the subspace associated with $x_1$ (i.e., events $e_1 \cap x_1$ and $e_0 \cap x_1$) and the subspace associated with $x_0$ (i.e., events $e_1 \cap x_0$ and $e_0 \cap x_0$). The matrix $R_3$ rotates the 2-dimensional subspaces defined by whether or not the feature values of the cause and effect match. That is, the subspace associate with matching feature values (i.e., events $e_1 \cap x_1$ and $e_0 \cap x_0$) and the subspace associated with mismatching feature values (i.e., events $e_1 \cap x_0$ and $e_0 \cap x_1$).

As before, we assume there is a unit length vector $|\psi\rangle$ in the vector space representing an individual's beliefs about the different feature combinations with coordinates $\alpha_{11}$, $\alpha_{10}$, $\alpha_{01}$, and $\alpha_{00}$ where $\alpha_{00} = \sqrt{1-(\alpha_{11}^2+\alpha_{10}^2+\alpha_{01}^2)}$ in order to ensure the state vector is unit length. In total, the 4-dimensional model has six parameters: $\theta_1$, $\theta_2$, $\theta_3$, $\alpha_{11}$, $\alpha_{10}$, and $\alpha_{01}$. As in the 2-dimensional model, probabilities are calculated by projecting the state vector onto different subspaces.

### 8-dimensional model

The 8-dimensional model assumes that individuals form mental representations of all joint events. Each of the eight dimensions is associated with one of the possible eight joint events such as $(e_1 \cap x_1 \cap y_0)$ and $(e_0 \cap x_0 \cap y_1)$. This is equivalent to a classical probability model with a single sample space describing all of the events. That is, joint probabilities as calculated using classical probability theory are equal to the probabilities that are calculated geometrically by projecting the state vector onto subspaces. Because all joint events exist, there are no rotation matrices. The model simply assumes there exists an 8-dimensional state vector where probabilities are calculated by projecting it onto different subspaces.

The state vector is composed of eight probability amplitudes $\alpha_{111}$, $\alpha_{110}$, $\alpha_{101}$, $\alpha_{100}$, $\alpha_{011}$, $\alpha_{010}$, $\alpha_{001}$, and $\alpha_{000}$ where the indices represent the feature values of $e$, $x$, and $y$ respectively. There are seven degrees of freedom in the model since the state vector must be unit length. The degrees of freedom can be reduced by adopting specific parameterizations. For example, we can parameterize the model in accordance with causal power.

In causal power, each cause $i$ is associated with a power parameter $w_i$ capturing the power of the cause to produce the effect. For experiment 1, there are two power parameters, $w_x$ and $w_y$, for the two causes of $e$. Cheng (1997) also assumed there could be alternative causes for the effect which might be known or unknown. These alternative causes are also associated with a power parameter labeled $w_a$. By the axioms of classical probability theory and independence of $x$ and $y$ we can write the joint probabilities for the three features as $p(e_i,x_j,y_k) = p(e_i|x_j,y_k)p(x_j)p(y_k)$ where $i$, $j$, and $k \in \{0,1\}$. Causal power theory assumes the conditional probability of the effect given the causes is computed using a "noisy-or" equation: $p(e_1|x_j,y_k) = 1-(1-w_x)^j(1-w_y)^k(1-w_a)$. Thus, five parameters are needed to define all eight joint probabilities in experiment 1: the three power parameters, $w_x$, $w_y$, and $w_a$, and the prior probabilities of the causes, $p(x_1)$ and $p(y_1)$. The eight joint probabilities can then be mapped directly to squared probability amplitudes. For example, $\alpha_{111}^2 = [1-(1-w_x)(1-w_y)(1-w_a)]p(x_1)p(y_1)$. As in the 2 and 4-dimensional models, probabilities are calculated by projecting the state vector onto different subspaces.

### Model Fits to Experiment 1

We fit the three models to the mean judgments for the 40 inference questions listed in table 1 by minimizing the sum of squared error (SSE) between each model and data. We then compared the models by using BIC values, which can be converted into Bayes factors by $e^{BIC_i - BIC_j}$ (Kass & Raftery, 1995). Using this approximation of the Bayes factor, the 2-dimensional model is strongly preferred to the 4-dimensional model and very strongly preferred to the 8-dimensional model (Kass & Raftery, 1995). Figure 2 compares model predictions to the observed data using the best fit parameters.

## Predictions from the 2-D Model

In fitting the data from experiment 1, the 2-dimensional model outperformed the 4 and 8 dimensional models with respect to BIC. However, a good model should not only fit observed data, but also generate new testable predictions. The 2-dimensional model makes two parameter free predictions about invariances. The first is called reciprocity (also known as the inverse fallacy (Villejoubert & Mandel, 2002)) and occurs when the probability of the effect given a cause is the same as the probability of the cause given the effect, e.g. $p(e_1|x_1) = p(x_1|e_1)$. The second prediction is a memoryless effect where the probability of a feature only depends on the most recent information given. For example, the memoryless property implies $p(x_1|y_1) = p(x_1|e_1,y_1)$ since $y_1$ is the most recent given information. We developed a new experiment to test both predictions.

### Methods

58 University of California, Irvine undergraduates participated in the experiment online for course credit. It was identical to experiment 1 except we used a within subjects design and asked participants to judge the probabilities listed in table 2. Participants judged each probability twice in two different randomized blocks.

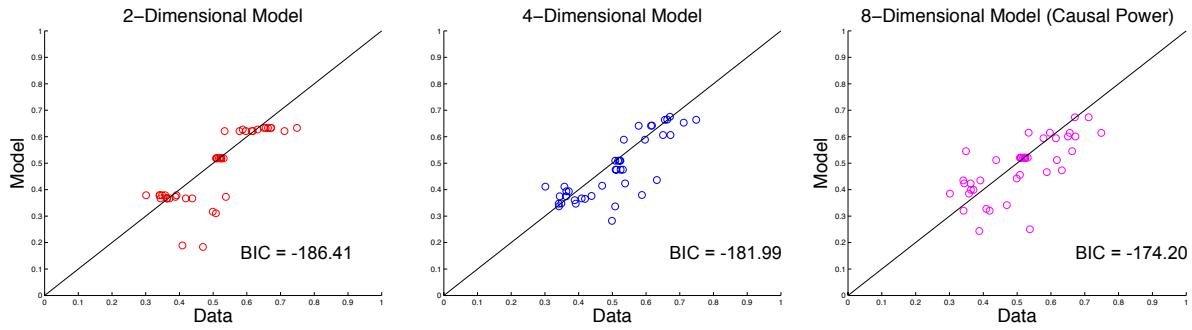


Figure 2: Predicted values compared to observed data from experiment 1 for three models. The 2-D model has the lowest BIC.

Table 2: Judgments in Experiment 2.

| | Probabilities | | Bayes Factor |
|---|---|---|---|
| reciprocity | $p(x_1 \mid e_1)$ | $p(e_1 \mid x_1)$ | 5.45 |
| | $p(y_1 \mid e_1)$ | $p(e_1 \mid y_1)$ | 6.59 |
| memoryless | $p(x_1 \mid y_1)$ | $p(x_1 \mid e_1, y_1)$ | 2.40 |
| | $p(x_1 \mid y_0)$ | $p(x_1 \mid e_1, y_0)$ | 1.68 |
| | $p(y_1 \mid x_1)$ | $p(y_1 \mid e_1, x_1)$ | 1.65 |
| | $p(y_1 \mid x_0)$ | $p(y_1 \mid e_1, x_0)$ | 6.93 |

## Results

The average score on the 10 question multiple choice test was 7.8. We excluded 17 participants with scores less than 7 out of ten. For our analyses, we averaged each individual's judgments from the two blocks. Because invariances correspond to null hypotheses and it is not possible to state evidence for the null hypothesis in standard significance testing, we calculated the Bayes factor for each comparison in table 2 following Rouder, Speckman, Sun, Morey, and Iverson (2009). Bayes factors greater than 1 tend to favor the null hypothesis and values greater than 3 are considered positive evidence for the null hypothesis (Kass & Raftery, 1995). In general, the predictions from the 2-dimensional model are supported.

## Discussion

The 2-dimensional model predicts that individuals break up complex reasoning problems into several simpler pieces. Judgments are made by sequentially processing these individual pieces. In causal situations with vague and imperfect information, such as the experiments discussed in this paper, a mental representation without joint events might be a simpler and more efficient way to evaluate events. Perhaps in situations where individuals have the opportunity to learn about casual relationships or are given statistical descriptions of these relationships, they can build more complete mental representations which include more joint events.

## Acknowledgments

JST was supported by NSF grant SES-1326275. EMP was supported by Leverhulme Trust grant RPG-2013-00 and AFOSR, Air Force Material Command, USAF, grant FA 8655-13-1-3044. The U.S Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright notation thereon.

## References

Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, *104*, 367-405.

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2011). Asymmetries in predictive and diagnostic reasoning. *Journal of Experimental Psychology: General*, *140 (2)*, 168-185.

Fernbach, P. M., & Sloman, S. A. (2009). Causal learning with local computations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 678-693.

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

Jenkins, H. M., & Ward, W. C. (1965). Judgment of contingency between responses and outcomes. *Psychological Monographs: General and Applied*, *79*, 1-17.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, *90*(430), 773-795.

Rehder, B. (2003). Categorization as causal reasoning. *Cognitive Science*, *27*, 709-748.

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t-tests for accepting and rejecting the null hypothesis. *Psychonomic bulletin & review*, *16*(2), 225–237.

Shanks, D. R. (1991). On similarities between causal judgments in experienced and described situations. *Psychological Science*, *2*(5), 341–350.

Sloman, S. A., & Fernbach, P. M. (2011). Human representation and reasoning about complex causal systems. *Information, Knowledge, Systems Management*, *10*, 1-15.

Trueblood, J. S., & Busemeyer, J. R. (2011). A quantum probability account of order effects in inference. *Cognitive Science*, *35*, 1518-1552.

Trueblood, J. S., & Busemeyer, J. R. (2012). A quantum probability model of causal reasoning. *Frontiers in Cognitive Science*, *3*, 1-13.

Villejoubert, G., & Mandel, D. R. (2002). The inverse fallacy: An account of deviations from bayes's theorem and the additivity principle. *Memory & Cognition*, *30*(2), 171–178.