When are representations of causal events quantum versus classical?

James M. Yearsley (james.m.yearsley@vanderbilt.edu)
Jennifer S. Trueblood (jennifer.s.trueblood@vanderbilt.edu)
Department of Psychology, Vanderbilt University, Nashville TN, USA

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

Abstract

Throughout our lives, we are faced with a variety of causal reasoning problems. Arguably, the most successful models of causal reasoning, Causal Graphical Models (CGMs), perform well in some situations, but there is considerable variation in how well they are able to account for data, both across scenarios and between individuals. We propose a model of causal reasoning based on quantum probability (QP) theory that accounts for behavior in situations where CGMs fail. Whether QP or classical models are appropriate depends on the representation of events constructed by the reasoner. We describe an experiment that suggests the representation of events can change with experience to become more classical, and that the representation constructed can vary between individuals, in a way that correlates with a simple measure of cognitive ability, The Cognitive Reflection Task.

Keywords: Causal reasoning, quantum probability theory, Bayes networks, order effects, Bayesian parameter estimation

Introduction

Reasoning about the causal relationships between events is an important component of everyday cognition that allows us to make sense of the world. For example, I know that when I plug in and turn on an electric kettle, water boils. Causal reasoning is also an essential component of problem solving, for example if I plug in and turn on the same kettle and water doesn't boil, I might infer that the fuse has blown.

Human reasoners are often very competent at performing causal reasoning tasks, in the sense of providing judgments aligning well with normative prescription. Probably the most successful class of models of causal reasoning are Causal Graphical Models (CGMs) based on Bayes nets. These models are normative, since they represent causal relationships using Bayes' calculus (Pearl, 1988) and have been successfully applied to a variety of human causal reasoning problems (Tenenbaum, Griffiths & Kemp, 2006; Griffiths & Tenenmaum, 2009). In spite of the apparent success of CGMs, there have been several recent studies that report violations of some of the predictions of these models including asymmetries between predictive and diagnostic reasoning, order effects, and violations of the casual Markov condition (Sloman & Fernbach, 2011; Trueblood & Busemeyer, 2012; Rehder, 2014). Equally, there seems to be considerable variation between individuals in how closely performance matches prescription (Rehder, 2014).

The challenge of modeling behavior that violates the normative prescriptions of classical (Bayesian) probability theory is not unique to causal reasoning. Recently, some researchers have been pursuing quantum probability theory (QP) as a way of modeling such behavior (Busemeyer & Bruza, 2011). QP is essentially the mathematical theory of probability associated with quantum theory, abstracted from the physics. QP contains features, such as contextuality, order effects and constructive judgments that appear to match well the way human decision makers often reason. One aim of this paper is to demonstrate that it is possible to build models of causal reasoning based on QP which provide a good description of behavior in situations where classical models fail.

However the introduction of QP models presents us with a new question, in what situations do people adopt a quantum representation of information as opposed to a classical one? We argue that classical and QP models can be seen as different cases in a hierarchy of models which differ in terms of the way events are represented. The QP model we introduce is in some sense the simplest possible, as it involves representing events in the smallest possible sample space (2D). Experience with a scenario may allow reasoners to construct a more complex representation, with a larger dimensional sample space, ultimately resulting in a fully classical model of causal relations. Equally, it is possible that simpler representations may be used to make heuristic judgments, while more complex representations are only formed when reasoning in a more deliberative way. If this is correct, the type of representation used by a decision maker may be linked to other individual differences such as the score on a Cognitive Reflection Test (CRT) (Frederick, 2005), or may be influenced by experimental instructions, such as speed vs accuracy prompts. Thus a second aim of this paper is to show that differences in performance in causal reasoning tasks can be explained in terms of the properties of the mental representation the individual constructs, and that this can be related both to experience with the task and to individual differences in cognitive ability.

We test our ideas in an experiment where participants were asked to make judgments involving conditional probabilities about various novel categories. We constructed a classical model and a QP model and used Bayesian analysis to compare them. Three main findings emerge, 1) Both the classical and the QP models provide good fits for some sets of participants, 2) The fits of the classical/QP models get better/worse as participants gain experience in the task, 3) Participants who scored highest/lowest on the CRT task tended to be better fit by the classical/QP model.

These findings suggest that the type of representation an individual constructs to reason about causal relations can vary both across individuals and with experience. We argue that

this may shed light on why classical Bayesian models of causal reasoning seem to match behavior in some situations, but not in others.

Experiment 1

Our experiment uses a paradigm developed by Rehder (2003) to study causal reasoning with novel categories. Our aim is to assess two phenomena from the judgment literature, order effects and the "inverse fallacy" (Villejoubert & Mandel, 2002), in causal reasoning and how these phenomena are related to individual differences in cognitive ability. Order effects occur when the order of information influences judgments (e.g., $p(E|X,Y)$ is not judged equal to $p(E|Y,X)$). It has also been shown that people commit what is known as the "inverse fallacy", that is, judging $p(X|Y) = p(Y|X)$. Both of these empirical findings are difficult to reconcile with an approach to causal reasoning based on Bayesian probability theory.

Methods

60 undergraduate students from Vanderbilt University participated in the experiment online at a time of their choosing for course credit. Participants were randomly assigned to one of two novel animal categories (either Lake Victoria Shrimp or Kehoe Ants). Each animal had three binary features (X, Y, and E) where two of the features (X, Y) causally influenced the third (E) to form a common effect network. For example, in the Lake Victoria Shrimp category, X = ACh neurotransmitter (high or low amount), Y = sleep cycle (accelerated or normal), and E = body weight (high or low). Participants were given information about the typicality of feature values. For example, they were told that "Most Lake Victoria Shrimp (90%) have a high amount of ACh whereas a few (10%) a low amount of ACh". In both categories, 90% of animals had feature X_1, 10% had feature Y_1, and 50% had feature E_1. After studying this information, they 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.

Participants were then told the causal relationships between features. These relationships were described as one feature causing another. In both categories, X_1 and Y_1 cause E_1. Participants were also told there were no known relationships between X and Y. After reading this information participants 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 relations.

After this test, participants completed six blocks of trials where they were asked to make judgments about the value of different features. There were two block types (BX and BY) that were repeated three times in an alternating fashion (e.g., BX, BY, BX, BY, BX, BY). Participants were randomly assigned to start with either the BX or BY block.

Each block contained nine judgment questions where participants were asked to select the value of a particular feature (see Table 1). At the start of each question, participants were told that a biologist caught a new animal (either shrimp or ant) and were queried about one of the features of that animal. For example, in the Lake Victoria Shrimp category, they might be asked "What type of body weight do you think this shrimp has?" Participants were given three response options: feature value 1, feature value 2, or equally likely to be feature value 1 or 2. For example, in the question about body weight, the response options were 1) a low body weight, 2) a high body weight, and 3) equally likely to be low or high.

Some questions asked participants to make a sequence of judgments about a feature value (e.g., E) as they learned new information about the other features (e.g., X, Y). For example, they might be asked about the body weight of a shrimp given lab tests that showed the shrimp had a high amount of ACh neurotransmitter (i.e., $E|X_1$). Participants might then be asked to reevaluate body weight based on additional lab tests that showed the shrimp also had a normal sleep cycle (e.g., $E|X_1,Y_2$). Note that information about the value of the first feature (e.g., X_1) remained on the computer screen when new information about the second feature (e.g., Y_2) was presented. This was to reduce the chance of memory failures. In the BX (BY) block, information about feature X (Y) was always presented before information about feature Y (X) in sequences involving both features. This helped reduce the influence of memory on future judgments about reverse orderings.

After finishing the six judgment blocks, participants completed the CRT (Frederick, 2005). This test assesses individual's ability to suppress a spontaneous and intuitive ("System 1") wrong answer in favor of a deliberative and reflective ("System 2") correct answer. The test consists of three items using a free-response format and is scored by counting the number of correct responses across the items. The CRT has been correlated with many behavioral measures including temporal discounting, mental heuristics, and risk preferences (Frederick, 2005; Toplak, West, & Stanovich, 2011).

Results

The average score on the 10 question multiple choice test was 9.6 indicating most participants correctly learned the feature values and causal relationships during the first part of the experiment. For analyses of the judgment data, we scored responses in a similar way to Rehder (2014) by assigning the following values to the three response options: feature value 1 = 1, feature value 2 = 0, and equally likely = 0.5. Note that there were no differences between judgments in the two different animal categories and so responses were collapsed for the following analyses. We also grouped individuals into three groups based on CRT scores: high = CRT score of 3, medium = CRT score of 1,2, and low = CRT score of 0. We combined individuals with CRT scores of 1 and 2 into a sin-

Table 1: Judgments in Experiment 1

Block	Judgments									
BX	E	X	Y	$E\|X_1$	$E\|X_1,Y_2$	$E\|X_2$	$E\|X_2,Y_1$	$X\|Y_2$	$Y\|X_1$
BY	E	X	Y	$E\|Y_1$	$E\|Y_1,X_2$	$E\|Y_2$	$E\|Y_2,X_1$	$Y\|X_2$	$X\|Y_1$

gle group so we had roughly an equal number of individuals per group. There were 21 participants in the CRT high group, 19 in the CRT medium group, and 20 in the CRT low group.

Order effects were assessed by comparing the judgments $E|X_1,Y_2$ and $E|Y_2,X_1$ and the judgments $E|X_2,Y_1$ and $E|Y_1,X_2$. For each individual, we calculated an "order effect score" defined as $|(E|X_1,Y_2)-(E|Y_2,X_1)|+|(E|X_2,Y_1)-(E|Y_1,X_2)|$. Higher scores indicate larger order effects and consequently a more "quantum" representation of information. By grouping the blocks into pairs (i.e., blocks 1 and 2, blocks 3 and 4, blocks 5 and 6), we can calculate three different order effect scores for each individual. These can be used to examine changes in order effects due to experience gained through repetition. The top panel of Figure 1 shows the order effect scores for the three CRT groups across the block pairs. A repeated measures ANOVA showed a main effect of order effect score ($F(2,118) = 9.86$, $p < .001$) and CRT group ($F(2, 57) = 3.45$, $p = 0.04$). A Bayesian analysis revealed a similar conclusion with a Bayes Factor of 405.34 for a model including order effect score and CRT group over a null model.

The inverse fallacy was assessed by comparing the judgments $X|Y_1$ and $Y|X_1$ and the judgments $X|Y_2$ and $Y|X_2$. Similar to the order effect score, we can also calculate an "inverse fallacy score" defined as $|(X|Y_1)-(Y|X_1)|+|(X|Y_2)-(Y|X_2)|$. Lower scores (closer to zero) indicate larger degrees of the fallacy and consequently a more "quantum" representation of information. Like the order effect score, we grouped blocks into pairs and calculated three different order effect scores for each individual. The bottom panel of Figure 1 shows the inverse fallacy scores for the three CRT groups across the block pairs. A repeated measures ANOVA showed a significant interaction between inverse fallacy score and CRT group ($F(4, 118) = 3.22$, $p = 0.015$). A Bayesian analysis revealed a similar conclusion with a Bayes Factor of 3.13 for a model including the interaction of inverse fallacy score and CRT group over a null model.

Conclusions

The results show a clear separation between the order effect scores for the high CRT group versus the other two, suggesting the high CRT group is more likely to be using a classical representation. Figure 1 also shows that as participants gain experience with the scenarios the degree to which they display the (non-classical) order effect decreases. This suggests that participants' representation of the events changes with experience; becoming gradually more classical and less quantum. The data also show an interaction between inverse fallacy score and CRT group. This provides evidence that the low CRT group is more likely using a QP representation. Below we directly compare the performance of a classical and

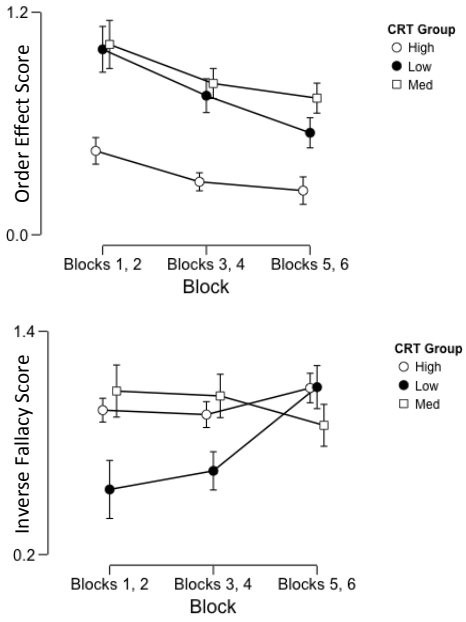

Figure 1: Top panel: Order effect scores for three CRT groups across block pairs. Bottom panel: Inverse fallacy scores for three CRT groups across block pairs. Error bars show the standard error.

QP model for this data.

Quantum Models

Before we introduce the details of the models, we provide a short introduction to QP theory. Recently a number of researchers have been investigating models of cognition based on the mathematics of QP theory (Busemeyer & Bruza, 2011). These models share the property that they are probabilistic, but the rules for assigning probabilities to events are those abstracted from quantum theory, rather than the usual Kolmogorov axioms. These models have features, such as contextuality, order effects and interference effects that seem to align well with human reasoning, at least in some cases.

One important topic in the study of quantum models of cognition is understanding the conditions under which they do and do not apply. Clearly much human decision making can be adequately described using Bayesian probability theory and in these situations, although quantum models can perfectly account for behavior also, they are clearly superfluous. The difference between classical and quantum models is often phrased in terms of the way different events are represented by a reasoner, either as *compatible* or *incompatible* (we will explain these terms shortly), and so attempts to delimit the realm of application of quantum models have often focussed on the question of whether different events may be

represented in a compatible way. There are few concrete results in this area, but it is generally believed that experience with a particular situation, either from previous familiarity or acquired through learning, may allow events to be represented in a compatible way, whereas relatively novel situations are more likely to be represented in a incompatible way. In addition, quantum models are invoked to explain a similar set of phenomena as heuristics (Busemeyer et al. 2011), and so it seems plausible that incompatible representations of events, associated with quantum models, should be preferentially used for decisions executed more quickly with little conscious deliberation.

Compatible events are ones that may be assigned a simultaneous truth value. Thus, if event X and event Y are compatible, the conjunction $X \wedge Y$ is well defined. Probabilities for compatible events obey the Kolmogorov axioms. Two immediate consequences are that for compatible events X and Y,

$$\begin{aligned} p(X \wedge Y) &= p(Y \wedge X), \\ p(X|Y) &= p(Y|X)\frac{p(X)}{p(Y)} \end{aligned} \tag{1}$$

(Note the connection to order effects and the inverse fallacy.)

Almost all events that we counter in everyday life can in principle be represented in a compatible way. However doing so requires that decision makers have access to the joint probabilities of all of these events. This may be unfeasible from the point of view of memory capacity, since the number of probabilities grows exponentially with the number of events being considered. Equally, these probabilities might be difficult to learn, since joint probabilities correspond to subsets of the sample space, and if it takes a finite number of previous experiences to learn the approximate measure of each subset, then the amount of experience required again grows exponentially with the number of events considered.

In contrast with compatible events, incompatible ones are those for which $X \wedge Y$ is undefined. Thus although the probabilities $p^Q(X)$ and $p^Q(Y)$ exist, the joint $p^Q(X \wedge Y)$ may not. Typically one can define a modified version of conjunction with an explicit ordering, eg $X \wedge Y$ is taken to mean X *and then* Y for incompatible variables. This implies that,

$$\begin{aligned} p^Q(X \wedge Y) &\neq p^Q(Y \wedge X), \\ p^Q(X|Y) &\neq p^Q(Y|X)\frac{p^Q(X)}{p^Q(Y)}. \end{aligned} \tag{2}$$

In quantum models, one can choose to model two events as either compatible or incompatible, depending on the representation one chooses. If all events are chosen to be compatible one recovers a classical model, while if no two events are compatible (except for the trivial case of an event and its negation) then one has a maximally quantum model. If there are more than two possible events then there can be intermediate representations where some subset of events are compatible. Thus quantum models encompass classical ones, and we should more accurately speak of a hierarchy of different representations, from fully quantum to fully classical.

In quantum theory events are represented by subspaces of a Hilbert space (essentially a vector space), with associated projection operators P_i, and the initial knowledge state by an operator on this space known as the density operator ρ. In a particular basis these are both simply matrices, and the computations involved in computing probabilities just linear algebra. The probability a participant assigns to event X given that her initial knowledge state is ρ is simply,

$$p(X) = \mathrm{Tr}(P_X \rho) \tag{3}$$

where Tr denotes the trace of a matrix. The probability a participant assigns to E given X is,

$$p(E|X) = \frac{\mathrm{Tr}(P_E P_X \rho P_X)}{\mathrm{Tr}(P_X \rho)} \tag{4}$$

The number of parameters it takes to specify ρ and the P_i's depends on the dimension of the Hilbert space, which in turn depends on whether the representation of different events is chosen to be compatible or incompatible. We will see in the next section how this works for the specific cases we are interested in, but generally there are some parameters that fix the initial state, and then some that determine the relationship between the projection operators representing different events (i.e., exactly how incompatible two events are).

The Classical and QP Causal Reasoning Models

In this section we briefly outline the QP and classical models of causal reasoning we propose to test. As mentioned above, the distinction between quantum and classical here is rather artificial - for our purposes a classical model is simply a QP one where all the events in question are compatible. Our models incorporate three possible events, two causes X and Y, and an effect E together with their negations. We are interested in probabilities for single events such as $p(X)$ and also conditional probabilities for the effect given one or more event, $p(E|X), p(E|X,Y)$ and for one cause given another, $p(X|Y)$.

The first choice we need to make is which events are compatible, and which incompatible. Because there are three events there are several possible options, for the present study we examine two choices; either all events are compatible, leading to a classical model, or all events are incompatible, leading to a maximally quantum model. Intermediate models exist, and may be desirable since the different models display different combinations of non-classical behaviors.

Let us consider the classical model first. Since all events are compatible it must be possible to assign truth values to propositions such as $X \wedge Y \wedge E$, and so the space needs to contain vectors which represent these states. Therefore our space needs to be 8D to encompass the set of possible states. We may use the following basis of states[1],

$$\begin{aligned} |XYE\rangle &= \begin{pmatrix}1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\end{pmatrix}^T, \\ |XY\bar{E}\rangle &= \begin{pmatrix}0 & 1 & 0 & 0 & 0 & 0 & 0 & 0\end{pmatrix}^T, \\ &\vdots \\ |\bar{X}\bar{Y}\bar{E}\rangle &= \begin{pmatrix}0 & 0 & 0 & 0 & 0 & 0 & 0 & 1\end{pmatrix}^T \end{aligned} \tag{5}$$

[1] We use standard Bra-Ket notation (Busemeyer & Bruza, 2011).

and projection operators,

$$\begin{aligned} P_X &= |XYE\rangle\langle XYE| + |XY\bar{E}\rangle\langle XY\bar{E}| \\ &\quad + |X\bar{Y}E\rangle\langle X\bar{Y}E| + |X\bar{Y}\bar{E}\rangle\langle X\bar{Y}\bar{E}| \\ &= \text{diag}(1,1,1,1,0,0,0,0) \text{ etc.} \end{aligned} \tag{6}$$

The initial state may be a general density matrix, however it turns out that the probabilities we compute are sensitive only to the diagonal elements of ρ. Therefore we may take,

$$\rho = \text{diag}(\rho_{11}, \rho_{22}, \ldots, 1 - \rho_{11} - \rho_{22} - \ldots \rho_{77}) \tag{7}$$

It is easy to compute the various probabilities of interest in terms of the ρ_{ii}. There are therefore seven parameters in the classical model, $\{\rho_{11}, \rho_{22}, \ldots, \rho_{77}\}$. Since all events in the classical model are compatible, there are no predicted order effects, e.g. we expect,

$$p(E|X,Y) = p(E|Y,X) \tag{8}$$

Now let us turn to the maximally quantum model. Since all events are incompatible we can span the space with any single pair $\{X, \bar{X}\}$. For this reason the space we need is 2D. Any two events such as X and Y are related to each other via a unitary transformation in this space. Any unitary transformation may be parameterised in the following way,

$$R_a = \begin{pmatrix} \cos(\theta_a) & -\sin(\theta_a)e^{i\phi_a} \\ \sin(\theta_a)e^{-i\phi_a} & \cos(\theta_a) \end{pmatrix} \tag{9}$$

and we have

$$P_X = R_X \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} R_X^\dagger \tag{10}$$

We will take the initial state to be diagonal[2], $\rho = \text{diag}(\rho, 1-\rho)$ and it turns out we can set one of the ϕ parameters to be 0 without loss of generality. Thus we have six parameters in total for the quantum model, $\{\rho, \theta_E, \theta_X, \phi_X, \theta_Y, \phi_Y\}$.

One interesting feature of the 2D model is that because all the events are represented by projection operators onto one dimensional subspaces, various expressions for the probabilities simplify. One example is,

$$\begin{aligned} p(X|Y) &= \frac{\text{Tr}(P_X P_Y \rho P_Y)}{\text{Tr}(P_Y \rho)} = \frac{\langle X|Y\rangle\langle Y|\rho|Y\rangle\langle Y|X\rangle}{\langle Y|\rho|Y\rangle} \\ &= |\langle X|Y\rangle|^2 = \frac{\langle Y|X\rangle\langle X|\rho|X\rangle\langle X|Y\rangle}{\langle X|\rho|X\rangle} = p(Y|X) \end{aligned} \tag{11}$$

which means that the 2D model exhibits the "inverse fallacy".

The causal reasoning models we have developed output probabilities for various combinations of events. However the experimental set up we used involves a choice rather than a judgment similar to experiments by Rehder (2014). For this reason we run each predicted probability through a softmax function to simulate the fact that participants are forced to choose between the possible alternatives rather than outputting the exact probability. The softmax function has a standard form (eg Rehder, 2014) and involves two extra parameters, λ, τ that are the same in the classical and QP models.

[2]This is a useful trick. It is difficult to check whether a general matrix is an allowable density matrix, but writing the matrix in a basis where it is diagonal the test becomes trivial.

Model Fitting

Parameters for the classical and quantum model were estimated using a Bayesian analysis for each of the nine conditions (CRT={High, Medium, Low} $\times$ Block={1+2, 3+4, 5+6}.) For the classical model the priors for the ρ_{ii} were all taken to be uniform in the interval $[0,1]$, and then normalized to ensure $\sum_i \rho_{ii} = 1$. For the quantum model the prior for the ρ parameter was taken to be uniform in the interval $[0,1]$ and the priors for all angle parameters were taken to be $(\pi/2) \times \beta(2,2)$. Three MCMC chains were used to estimate the posterior distributions using JAGS. Chain convergence was assessed using the $\hat{R}$ statistic and all chains had good convergence behavior.

For each condition we also used JAGS to compute the DIC associated with the classical and QP models. The DIC is a generalization of the BIC, wth smaller values indicating a better model fit. A difference of three in the DIC between two models is usually taken to be indicate a significant difference in fit. The results are shown in Figure 2.

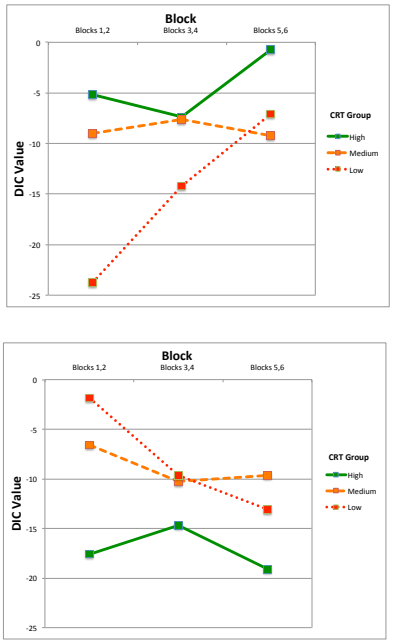

Figure 2: Top panel: DIC values for the QP model as a function of block number for each of the CRT groups. Bottom panel: DIC values for the CP model as a function of block number for each of the CRT groups.

For the QP model the DIC values show the following patterns; first the model performs better in participant groups where the CRT score is lower. Second, while the model performance for the medium and high CRT groups does not vary much across blocks, performance clearly decreases with block number for the low CRT group. For the classical model the opposite behavior is observed; first the model performs better in participant groups with a higher CRT score. Second, while performance for the high and medium CRT groups does not vary much across blocks, performance clearly increases

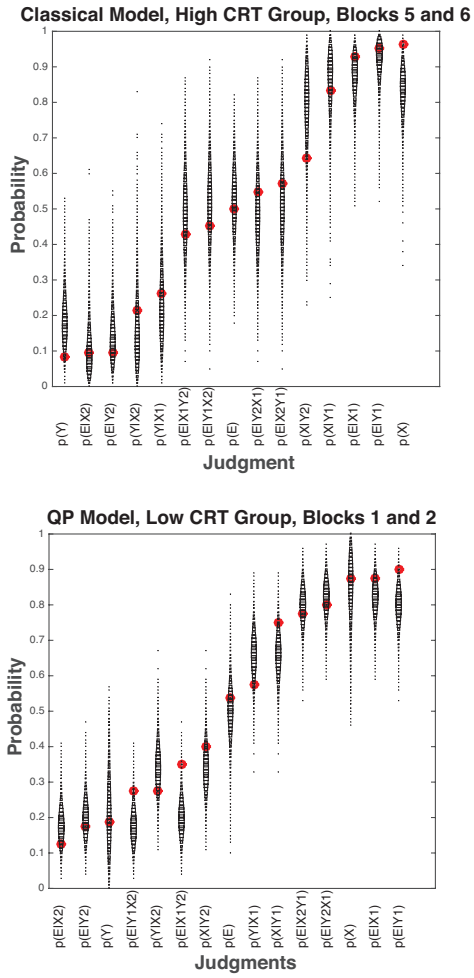

Figure 3: Top panel: Posterior probabilities (black squares) compared with empirical data (red dots) for the CP model, high CRT group and blocks 5+6. Bottom panel: Posterior probabilities compared with empirical data for the quantum model, low CRT group and blocks 1+2.

with block number for the low CRT group. Two examples of the model fits are shown in Figure 3.

Conclusion

The conclusions from the Bayesian model fitting corroborate the evidence of the diagnostic "order effect" and "inverse fallacy" scores. Overall, the representation of events that participants use to reason about causal relations appears to change as participants gain familiarity with the scenario, from an initially quantum or incompatible one to a final classical or compatible one. In addition, we found evidence that there are individual differences between participants in terms of their tendency to use classical or QP representations, which is associated with the CRT score.

Overall, our work sheds light on why causal graphical models (CGMs) have been successful in many situations, but can sometimes fail to agree with behavior. Equally it helps us understand why QP models can sometimes be successful but are superfluous in many cases. Reasoning about causal relations is neither inherently classical or quantum but rather is tied to the representation of events constructed by the reasoner. For novel scenarios or when reasoning quickly, representations of events may be incompatible and QP models are appropriate, however experience or more deliberative reasoning can lead to the formation of more complex compatible representations, which support classical causal reasoning.

Acknowledgments

JMY and JST were supported by NSF grant SES-1556415. EMP and JMY were supported by Leverhulme Trust grant RPG-2013-00. EMP was supported by Air Force Office of Scientific Research (AFOSR), Air Force Material Command, USAF, grants FA8655-13-1-3044. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright notation thereon.

References

Busemeyer, JR & Bruza, P (2011). Quantum models of cognition and decision. CUP: Cambridge, UK.

Busemeyer, JR, Pothos, EM, Franco, R, & Trueblood, J. (2011). A quantum theoretical explanation for probability judgment errors. *Psych. Rev.*, 118, 193-218.

Frederick, S (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives.* 19, 25-42.

Griffiths, TL & Tenenbaum, JB (2009). Theory-based causal induction. *Psych. Rev.* 116(4), 661-716.

Pearl, J (1988). Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufmann.

Rehder, B (2003). A causal-model theory of conceptual representation and categorization. *JEP:LMC* 29, 1141-1159.

Rehder, B (2014). Independence and dependence in human causal reasoning. Cognitive psychology, 72, 54-107.

Sloman, SA & Fernbach, PM (2011). Human representation and reasoning about complex causal systems, *Information, knowledge, systems management* 10, 1-15.

Tenenbaum, JB, Griffiths, TL & Kemp, C (2006). Theory-based Bayesian models of inductive learning and reasoning. *Trends in Cognitive Sciences*. 10(7), 309-318.

Toplak, ME, West, RF, Stanovich, KE (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*. 39, 1275-1289.

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

Villejoubert, G, & Mandel, DR (2002). The inverse fallacy: An account of deviations from Bayes' theorem and the additivity principle. Memory & Cognition, 30(2), 171-178.