RUNNING HEAD: Individual differences in causal reasoning

A quantum probability account of individual differences in causal reasoning

Percy K. Mistry
University of California, Irvine
pkmistry@uci.edu

Emmanuel M. Pothos
City University, London
emmanuel.pothos.1@city.ac.uk

Joachim Vandekerckhove
University of California, Irvine
joachim@uci.edu

Jennifer S. Trueblood
Vanderbilt University
jennifer.s.trueblood@vanderbilt.edu

Corresponding Authors:
Percy K. Mistry
Department of Cognitive Sciences,
2201 Social & Behavioral Sciences Gateway Building (SBSG),
University of California
Irvine, CA 92697-5100
phone: +1 949-220-3339
email: pkmistry@uci.edu

Jennifer S. Trueblood
Department of Psychology
Vanderbilt University
PMB 407817
2301 Vanderbilt Place
Nashville, TN 37240-7817
phone: 615-343-7554
email: jennifer.s.trueblood@vanderbilt.edu

Abstract

We use quantum probability (QP) theory to investigate individual differences in causal reasoning. By analyzing data sets from Rehder (2014) on comparative judgments, and from Rehder & Waldmann (2016) on absolute judgments, we show that a QP model can both account for individual differences in causal judgments, and why these judgments sometimes violate the properties of causal Bayes nets. We implement this and previously proposed models of causal reasoning (including classical probability models) within the same hierarchical Bayesian inferential framework to provide a detailed comparison between these models, including computing Bayes factors. Analysis of the inferred parameters of the QP model illustrates how these can be interpreted in terms of putative cognitive mechanisms of causal reasoning. Additionally, we implement a latent classification mechanism that identifies subcategories of reasoners based on properties of the inferred cognitive process, rather than post hoc clustering. The QP model also provides a parsimonious explanation for aggregate behavior, which alternatively can only be explained by a mixture of multiple existing models. Investigating individual differences through the lens of a QP model reveals simple but strong alternatives to existing explanations for the dichotomies often observed in how people make causal inferences. These alternative explanations arise from the cognitive interpretation of the parameters and structure of the quantum probability model.

Keywords: individual differences; causal reasoning; quantum probability; causal graphical models; Bayesian inference

1. Introduction[1]

Our ability for causal reasoning is arguably fundamental for many of the achievements that we consider uniquely human, science and engineering for example, as well as being an integral aspect of competence in many aspects of day-to-day life. Our causal reasoning ability demonstrates a staggering scope of applicability, from extremely simple situations (e.g., if the switch is turned off, the appliance will stop working) to extremely complex ones (e.g., the impact of fiscal policy changes on economic climate), involving different numbers of implicated variables, different structural relationships between the variables, and differences in the certainty of the information available for a causal reasoning problem. Moreover, there is evidence that reasoning ability may vary depending on whether judgments are being made from experience, statistically described contingencies, or linguistic narratives and descriptions (Shanks, 1991).

These observations perhaps paint a grim picture regarding the possibility of a general model of human causal reasoning, yet psychologists have been intensely engaged with this endeavor. Several influential models have been proposed, including ΔP (Jenkins & Ward, 1965) and power PC theory (Cheng, 1997), but important shortcomings (e.g. inability to account for a large range of experimental data, including among others, asymmetries between diagnostic and predictive inferences, the influence of uncertainty in diagnostic judgments, etc.) have been identified for such traditional approaches (Sloman & Fernbach, 2011; Trueblood & Busemeyer, 2012; Lober and Shanks, 2000; White, 2005; Fernbach, Darlow, & Sloman, 2010). Currently,

[1] Key Abbreviations used: quantum probability (QP), classical probability (CP), common cause network (CC), chain network (CH), common effect network (CE), causal graphical models (CGMs), conjunctive model (CONJ), associative random Markov field model (ASSC), specific shared disabler based model (DISAB), deviance information criterion (DIC).

one of the predominant modeling approaches for causal reasoning uses Causal Graphical Models (CGMs; e.g. Tenenbaum, Griffiths, and Kemp, 2006; Griffiths and Tenenbaum, 2009; Fernbach and Sloman, 2009; Goodman, Ullman, and Tenenbaum, 2011; Kemp, Goodman, and Tenenbaum, 2010; Pearl, 2014), which are a graphical way to represent causal relations based on Bayes nets (Kim & Pearl, 1983; Pearl, 1988). In a CGM, variables are represented by nodes where the links between nodes and the direction of such links represent causal relationships, with the originating node called a parent and the end node a child. The rules of classical probability (CP) theory are employed to relate probabilities between nodes in parent – child relationships. CGMs also assume the Markov property, whose intuitive meaning is that the conditional probabilities for a node depend only on its parents. More formally, the Markov property states that any node in a CGM is conditionally independent of its non-effects, given its direct causes (Russell & Norvig, 2003). The role of the Markov property is that it greatly simplifies conditional probability computations, especially in complex causal structures, in which the range of dependencies, left unchecked, can quickly become intractable.

CGMs have been a key development in the study of causal reasoning (Griffiths et al., 2010; Oaksford & Chater, 2009). In work separate to cognitive psychology, CGMs are often researched and employed in industry applications relating to complex reasoning situations (e.g. Cowell et al., 1999). They benefit from excellent descriptive success (Tenenbaum, Griffiths, and Kemp, 2006; Griffiths and Tenenbaum, 2009; Goodman, Ullman, and Tenenbaum, 2011; Kemp, Goodman, and Tenenbaum, 2010), and allow a clear statement of the principles that guide human causal reasoning competence.

The motivation for the present work is the increasing evidence that, while the descriptive performance of CGMs is excellent in many cases of human causal reasoning, there are consistent

violations of CGM principles as well. For example, there have been reports of violations of the Markov condition (Rottman & Hastie, 2014; Park & Sloman, 2013; Rehder, 2014; Fernbach & Sloman, 2009; Waldmann, Cheng, Hagmayer, & Blaisdell, 2008; Hagmayer & Waldmann, 2002; Rehder & Waldmann, 2016), as well as reports of findings that are inconsistent with the CP principles that CGMs adhere to, notably selectively neglecting alternate causes in predictive (but not diagnostic) judgments (Fernbach, Darlow, & Sloman, 2010), and failure to discount based on the presence of alternative causes (Rehder, 2014; Rehder & Waldmann, 2016). To illustrate, consider an individual reasoning about a friend's recent weight loss. Anti-discounting occurs when the individual thinks that it is more likely the friend improved their diet after learning that their friend lost weight and started exercising as compared to only learning about weight loss (assuming diet and exercise are independent). Mathematically, the probability of improved diet is higher when only weight loss is known since also knowing about a new exercise regime explains the weight loss. In this simple example, the individual is reasoning associatively about diet and exercise rather than reasoning according to the rules of probability theory.

How are we to approach the dual challenges of consistency between human behavior and CGM principles in many cases (not to mention the normative justification for CGMs, that is, the fact that CGMs represent causal relationships using Bayes' calculus and thus respect classical probabilistic norms, Kim & Pearl, 1983; Pearl, 1988) and the increasing evidence that, at least in some cases, naïve observers reason in a way that conflicts with CGM prescription? We clearly do not want to abandon CGMs entirely, but it is equally clear that human causal reasoning must reflect a *combination* of CGM principles and principles based on alternative mechanisms. Such alternative mechanisms could be non-normative modifications of CGMs, where deviation of reasoning from the normative process is explained by way of "cognitive shortcuts," such as the

strategic decision to neglect alternatives (Fernbach and Rehder, 2013), augmenting CGMs with additional hidden variables not part of the actual experimental situation (hidden mediators, disablers, enablers or causes, e.g. Rehder, 2014), or deviations from CP calculations (e.g., computing conditional probabilities as conjunctive probabilities or ignoring causal direction in conditionalization, Rehder, 2014).

Rehder (2014) provided an extensive investigation of causal reasoning, in relation to CGMs and three alternative non-normative reasoning strategies. To illustrate these strategies, we employ one of the examples originally used by Rehder (2014), where three events – high or low retirement savings, high or low trade deficits, and high or low interest rates – are causally linked in different ways. The first non-normative strategy is the Conjunctive Model (CONJ) and its characteristic is that conditional probabilities are evaluated conjunctively. For example, the probability of high retirement savings given low trade deficits would be instead evaluated as a joint probability, the probability of high retirement savings and low trade deficits. Note that, while there is evidence for a CONJ strategy in Rehder (2014), the model is clearly ad hoc. For example, in other probabilistic judgment scenarios (e.g., the conjunction fallacy), the opposite is sometimes assumed, that is, that the evaluation of conjunctive probabilities is computed as a function of conditional probabilities (Tenenbaum & Griffiths, 2001). The second strategy is the Specific Shared Disabler Model (DISAB). Per DISAB, a hidden disabling mechanism assumes an additional variable imagined by participants that probabilistically influences one or more of the existing causal mechanisms. For example, a participant might envisage an additional variable (not part of the experimental scenario), foreign exchange rates, which might probabilistically moderate the causal relationship between interest rates and trade deficits. Again, while there is evidence for a DISAB strategy in terms of its descriptive success, the choice of structure for the

hidden variable remains ad hoc. The third strategy is the Associative Model (ASSC), which posits an associative Markov random field. This essentially assumes an associative (correlational) relationship between the variables of interest, without allowing for any specific direction of causality.

Successful description of the aggregate causal reasoning data across Rehder's (2014) four experiments required a model that was a weighted linear combination of *all* four strategies (a normative CGM strategy and the three non-normative ones). Each of these models had between three to five free parameters, with an additional three free parameters for the mixture weighting of these models. So, Rehder (2014) can be taken as one of the most specific demonstrations that human causal reasoning embodies both a normative and a non-normative influence. Further, this mixture model accounted for results primarily at an aggregate level. The primary purpose of Rehder (2014) was to test for violations of the Markov principle and, based on a cluster analysis, identified two groups of participants, causal and associative (non-causal) reasoners. The latter group committed a higher number of Markov violations and exhibited greater anti-discounting behavior. The application of Rehder's (2014) analyses to the subgroups revealed a range of strategies for both subgroups, but with the expected biases towards the normative strategy (average weight of 0.59) for causal reasoners and the associative strategy (average weights of 0.67) for associative reasoners. Note, the associate strategy (ASSC) specifically refers to a lack of causal direction, but Rehder (2014) labelled associative reasoners as participants that mostly violated the classical CGM model (and whose behavior may be explained by one or more of the ASSC, CONJ, or DISAB models), but did not necessarily employ the ASSC strategy or display a complete insensitivity to causal direction. In our analysis, we continue to refer to associative

reasoners in this latter broader sense, as non-causal reasoners, whose behavior cannot be described by the classical CGM model.

For researchers interested in the principles underlying human causal reasoning, Rehder's (2014) analysis is groundbreaking, but it also raises three important questions. First, does the observed multiplicity of strategies primarily concern a within-participant description level or a between participants one? In other words, is it the case that different participants predominantly adopt a single strategy (so that averaged results require a model based on all four strategies) or is it the case that for each participant there are varying influences from all strategies? Rehder's classification of participants into associative and causal reasoners is suggestive, but ideally an individual differences analysis would be informed by parameters of the underlying cognitive process or processes and reveal directly the extent to which the postulated strategies are represented across most participants or few participants. Second, some of the non-normative strategies are ad hoc and have primarily descriptive value. Is it possible to propose a formalism that will encompass as special cases of its application both the normative and non-normative strategies? That is, can individual behavior patterns that have been conceptualized as involving qualitatively different causal models be accommodated through alterations of continuous parameters within a single framework, rather than as a mixture of ad hoc strategies? Third, does an application of a formal model for normative violations in causal reasoning enable predictions about new effects or insights in causal reasoning?

Making progress with these three questions leads to the two objectives of the present paper. First, we implement the relevant models as Bayesian Hierarchical models, so that model parameters are given hierarchical priors with hyper-parameters that allow us to systematically capture putative individual differences. The Bayesian models allow us to infer the posterior

distribution of the model parameters that best describe the observed data. This allows us to construct a posterior predictive distribution, which is the probability distribution over all possible data points given the posterior distribution of the parameters inferred, having seen the actual observed data. This is essential to understand exactly whether the diversity in causal reasoning strategies is within or between participants. Second, we examine a new model of causal reasoning based on quantum probability (QP) theory and use a Bayesian Hierarchical approach to explore whether the latent classification of individuals based on QP parameters (Lee & Webb, 2005) is consistent with the distinction between causal and associative reasoners that Rehder (2014) proposed. How important is this (intuitive) distinction in describing the data? A more rigorous individual differences approach may also allow novel perspectives in causal reasoning, as indeed Stanovich and West (2000) argued regarding the rationality debate.

Bayesian hierarchical models allow us to characterize any underlying cognitive model (both based on classical probability as well as QP models) in terms of basic parameters at an individual level, which are themselves specified as being generated by another process characterized by hyper-parameters. This allows capturing the nature of individual differences in behavior in a systematic and structured manner, providing simultaneous posterior distributions on individual and population level parameters. This approach can be used to specify multiple psychological processes (within a cognitive model) and perform a latent clustering of individuals depending on which process is the most likely to generate individual behavior. Finally, the approach can also be used to define a mixture model of different cognitive models, and provide a direct comparison between such models that may differ in terms of their underlying assumptions, including comparing classical probability and QP models within the same hierarchical

framework. A detailed perspective on the use of hierarchical Bayesian models for cognitive modeling can be obtained from Lee (2011).

A Bayesian hierarchical approach to individual differences carries a high computational burden, especially where multiple models are involved. This restricts the range of datasets that can be considered, compared to aggregate-level analyses. We chose to focus on two datasets, Rehder (2014) and Rehder and Waldmann (2016). Rehder's (2014) focus was exactly to test for violations of the Markov principle, which is consistent with one present objective, i.e., the (more) formal description of non-normative influences in causal reasoning (with QP; see shortly). Moreover, even though Rehder's (2014) main objective was not individual differences, his selection of inference problems led to evidence for three distinct non-normative strategies and corresponding evidence for individual differences. Thus, in seeking to understand individual differences using the present Bayesian Hierarchical approach, his dataset is highly suitable. Finally, the dataset was carefully constructed (315 participants, four well-controlled experiments manipulating content of the inference problem and causal structure, including common cause, chain, and common effect). Rehder (2014) tested for *relative* judgments between scenarios, that is, what combination of events under a causal network were more likely. To test the generalizability of our approach, we also examine a dataset from Rehder and Waldmann (2016), which uses *absolute* probability judgments rather than relative judgments.

The second objective of the paper is to explore whether the range of specific strategies, including the normative one that Rehder (2014) proposed, could be subsumed within a single formal model. We propose a model based on the principles of quantum probability, by which we mean the rules for how to compute probabilities, from quantum mechanics, without any of the physics. QP can lead to cognitive models very similar in nature to those using CP: in both cases,

the objective is a top-down or function first (Griffiths et al., 2010) description of cognition, emphasizing the computational principles that guide behavior (cf. Marr, 1982), but with limited process assumptions. QP is basically a framework for probabilistic inference alternative to CP. There is a motivation to consider QP for cognitive modeling, instead of CP, exactly for situations where human behavior appears at odds with the prescription from CP. In recent years, QP models of cognition have been successfully applied to various domains, including among others, decision-making (Pothos & Busemeyer, 2009), perception (Atmanspacher & Filk, 2010), probability judgments (Busemeyer, Wang, Pothos, & Trueblood, 2015; Trueblood & Busemeyer, 2011), similarity (Pothos, Busemeyer, & Trueblood, 2013; Pothos & Trueblood, 2015) memory (Brainerd et al., 2013), and conceptual categorization and knowledge formation (Aerts, Sozzo, & Veloz, 2016; Sozzo, 2015). Apart from understanding behavior that demonstrates violations of CP, such models have also been used to formalize cognitive notions such as psychological uncertainty, non-decomposability of cognition, a two layered structure of human reasoning including classical logical and quantum emergent processes, order effects and sensitivity to measurements (for overviews see Aerts, Broekaert, Gabora, & Sozzo, 2013; Aerts, Gabora, & Sozzo, 2013; Busemeyer & Bruza, 2012; Pothos & Busemeyer, 2013; Wang, Busemeyer, Atmanspacher & Pothos, 2013).

An interesting feature of QP is that representations can be *compatible* or *incompatible* (these are technical terms in QP). When representations are compatible, QP predictions are consistent with CP ones. By contrast, with incompatible representations we obtain many non-classical features in the computation of probabilities, for example, violations of the law of total probability. Psychologically, incompatibility means that the order in which events are processed critically affects behavior (thinking about one thing first influences how you think about the next

thing), similar to priming effects. When a situation involves multiple events, some events can be compatible and others can be incompatible. This results in a mixture of both "quantum-like" and "classical-like" properties. In typical causal reasoning situations, there are often several variables (causes/ effects) and so a hierarchy of causal reasoning models can be specified, from fully classical to fully quantum, depending on how many variables are pairwise incompatible (Trueblood, Yearsley, & Pothos, 2017).

Presently, we focus on the 'most quantum' possible QP model of causal reasoning, which treats all variables as incompatible. As will be discussed in the Quantum Probability Model section, the assumption of full incompatibility leads to an overall two-dimensional space, with all questions represented as rays. Because this model uses the lowest possible dimensional space (i.e., events in the experiments are binary and must be minimally represented using two dimensions), it offers a very simple account of the data. It is pertinent to focus on this simple QP approach, exactly because Rehder (2014) focused on inferences with a high expectation of non-classicality. If a simple QP model can account for the Rehder (2014) and Rehder & Waldmann (2016) results, this would be an important demonstration of the relevance of quantum principles in causal reasoning and, moreover, the (relative) mathematical simplicity of the model will facilitate the in-depth individual differences analyses. Note, we use exactly the same QP causal reasoning model for all of Rehder's (2014) experiments and for both comparative (Rehder, 2014) and absolute (Rehder & Waldmann, 2016) judgments.

In sum, Rehder (2014) proposed three heuristic strategies that would complement the normative CP one. Can Rehder's (2014) heuristic strategies be subsumed within a unified, (more) formal description within a simple QP model? At the level of individual differences, what are the properties of any classification based on latent (model) parameters and how consistent is

this with the associative vs. normative distinction Rehder (2014) reported? And finally does the simple QP model reveal new insights/ effects about causal reasoning? Finally, we note that the simplest, 'most quantum' QP model we will shortly present is unlikely to be a general model of causal reasoning – a more general approach is presented in Trueblood et al. (2017), though the additional complexity of this more general approach makes it unsuitable for an in-depth individual differences analyses. The motivation for employing such a model presently goes hand in hand with the specific focus of Rehder (2014) and Rehder & Waldmann (2016) on non-classicality in causal reasoning.

2. Description of Experiments and Results

We briefly describe the experiments and results from Rehder (2014) and Rehder & Waldmann (2016).

2.1. Comparative judgments (dataset 1, Rehder, 2014):

Task

In Rehder (2014), participants were taught one of the three causal network structures (common cause, chain or common effect) encompassing a set of relationships between three binary variables as shown in Figure 1. The causal networks were instantiated in either a domain-general (abstract) or domain-specific (economics, sociology, or meteorology) environment. Causal relationships between variables were described to the participants as independent causal processes. For example, in the economics domain, the variables were interest rates, trade deficits, and retirement savings, each of which could be large (high) or small (low); a relationship could take the form: low interest rates cause small trade deficits. The causal relationships taught to the participants were specified in a single sense, so that if a high (low) value of a cause facilitated

the presence of an effect, the low (high) value did not have the opposite effect (e.g. if low interest rates caused small trade deficits, high interest rates were causally unrelated to trade deficits). An example of a common effect structure is with interest rates and trade deficits both exerting a causal influence on retirement savings. Participants in this example were told that low interest rates cause high retirement savings and small trade deficits also cause high retirement savings. Each relationship was supported by a brief justification of how such a causal mechanism might work. For instance, the justification for low interest rates causing high retirement savings was given as "*Low interest rates stimulate economic growth, leading to greater prosperity overall, and allowing more money to be saved for retirement in particular*". Participants were then asked to make comparative judgments. On each trial, they were presented with two different situations (a situation is a particular combination of values for each node) in the causal structure, and asked to judge in which of these two situations was the target variable more likely to have a specific value (e.g., to have a low value). In the above example of the common effect structure, the two situations could be one where trade deficits were small and retirement savings were high, and the second could be one where trade deficits were unknown and retirement savings were high. The participants would then be asked to judge under which of these situations was a low value of interest rates more likely, that is, a comparative judgment. In this situation, participants should *normatively* believe that interest rates are more likely to be low in the second situation where trade deficits are unknown, rather than the first where a known alternative cause (low trade deficits) for the observed effect (high retirement savings) already exists.

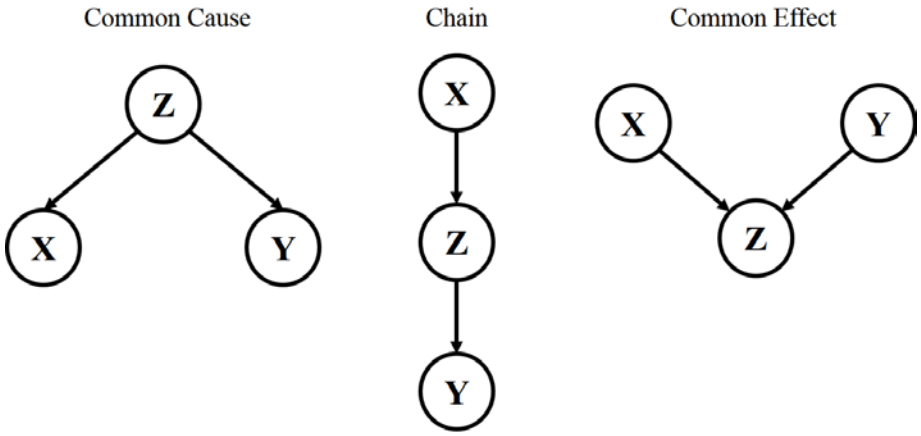


Fig. 1. Three types of causal network structures tested (adapted from Rehder, 2014). The circular nodes represent variables and the arrows depict the causal relationships.

Design

Rehder (2014) examined people's decisions in eight network states (situations A to H; see Table 1) that arise when evaluating the state of an unknown target variable considering all possible values of the remaining two variables (denoted X and Z), namely '0' (representing a state value that indicates the absence of a cause or effect), '1' (representing a state value that causally influences or is influenced), or '?' (representing an unknown value). For example, in the domain of economics if low interest rates caused small trade deficits, low interest rates and small trade deficits were coded as '1' and higher interest rates and high trade deficits as '0'. Note that the eight network states are not all possible combinations (3 possible states and 3 variables gives 27 possible network states). Of these, 3 include states where both the remaining variables are unknown, and were excluded. Of the remaining 24 states, 16 states were tested, however given the symmetry of the common cause and common effect structures, the inference for the two effects or two causes is similar (as was shown empirically), and the 16 different network states were collapsed to 8 unique states. The remaining 8 network states were not tested since they

were not expected to provide any insight as to whether normative CGM rules were being violated or whether discounting behavior was being exhibited.

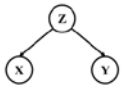
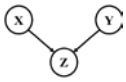
Participants were asked to compare the states A vs B, B vs C, D vs E, F vs G and G vs H, and indicate which of the two situations made the target variable (Y) more likely, or whether the variable was equally likely across the two situations. For our analyses, we combined the data from four experiments in Rehder (2014) (i.e., Experiments 2, 3, 4A and 4B). The different experiments tested different conditions, such as controlling for abstract versus concrete domains, specifying probabilistic causal relationships versus control conditions where no information on the strength of the causal links was provided and controlling for the base rate of questions where the two situations were equally likely. The analysis in the original paper suggested no significant differences based on these manipulations. Across these four experiments, there were 315 participants (105 per causal structure). Each participant made twenty such comparative judgments, with the causal structure (common cause, chain, common effect) and domain of variables (economics, sociology, meteorology, and an abstract domain in one condition) as between-subject conditions.

Normative expectations

Table 1 shows the normative predictions for each possible pair of situations based on a causal Bayes net model of the inference problem (see Rehder, 2014 for a detailed analysis of the normative predictions). Two key properties on which these predictions are based are the causal Markov property (also referred to as the parent-child property) and discounting in the common effect structure. To illustrate the Markov property, consider a common cause network where low interest rates cause both small trade deficits and high retirement savings. The causal Markov condition implies that if the value of interest rates is known, knowledge of trade deficits does not

provide any additional information towards the value of retirement savings, and vice versa. All information regarding the effect is captured in the node representing the common cause and the causal link between the two, if the value of the cause is known.

Table 1. Enumeration of the eight different situations in dataset 1 (A to H) for the common cause, chain and common effect networks, and data set 2 (A to E, I to K) for the common cause and common effect networks. Participants were required to make inferences about the '*target*' variable, given the states (0, 1 or unknown=?) of the remaining two variables. The normative relative predictions for situations A to H based on the causal Markov condition of causal graphical models and discounting in the case of common effect structures yield the predictions below (adapted from Rehder, 2014; Rehder & Waldmann, 2016). Situation J was tested twice, with each of the two causes (effects) set to 1 separately.

				Situation under which target is more likely		
Situation	X	Z	Y	Common cause 	Chain 	Common effect 
A	1	1	target			
B	?	1	target	A=B=C	A=B=C	C > B > A
C	0	1	target			
D	1	?	target			
E	0	?	target	D >> E	D >> E	D = E
F	1	0	target			
G	?	0	target	F=G=H	F=G=H	F=G=H
H	0	0	target			
I	1	target	1			
J_1	1	target	0	I > J > K	n/a	I > J > K
J_2	0	target	1			
K	0	target	0			

In a common effect structure, discounting refers to the phenomenon that the presence of a cause is deemed less likely when an alternate cause is present, than when no alternate cause is present. Take the example of low interest rates and small trade deficits both causing high retirement savings. Suppose that retirement savings are high. What can we say about the value of trade deficits? If we also know that interest rates are low, then the high retirement savings are plausibly a result of this cause, which makes the presence of small trade deficits redundant. That is, the presence of one cause is sufficient to explain the effect, making the alternate cause

redundant. Discounting refers to this latter inference, and is normatively expected behavior in a common effect structure.

Results

Rehder (2014) found that a significant number of participants violated one or both properties (causal Markov condition and discounting). About 23% of the 315 participants exhibited some form of reasoning that deviated from the predictions of CGMs, in particular, demonstrating a lack of sensitivity to causal direction (the associative reasoners). Insensitivity to causal direction can result in behavior that appears to ignore conditional independence as stipulated by the causal Markov property and exhibit anti-discounting behavior (i.e. judging the target cause as highly probable based on the presence of an alternative cause, which is opposite to normative expectation). But Rehder (2014) also identified several participants whose behavior more closely resembled the predictions of CGMs (the causal reasoners). Note, the participants labelled as associative and causal reasoners in Rehder (2014) both displayed multiple influences in their behavior. Figure 2 shows the normative (CGM) predictions and the actual observed aggregated mean choice responses for participants classified as causal and associative reasoners separately. The deviations from normative predictions for associative reasoners across all three networks are significant. Note that all behavioral patterns captured as associative reasoners, as well as the patterns observed under common effect structure for the cluster labeled causal reasoners required ad hoc explanations beyond the normative CGM model.

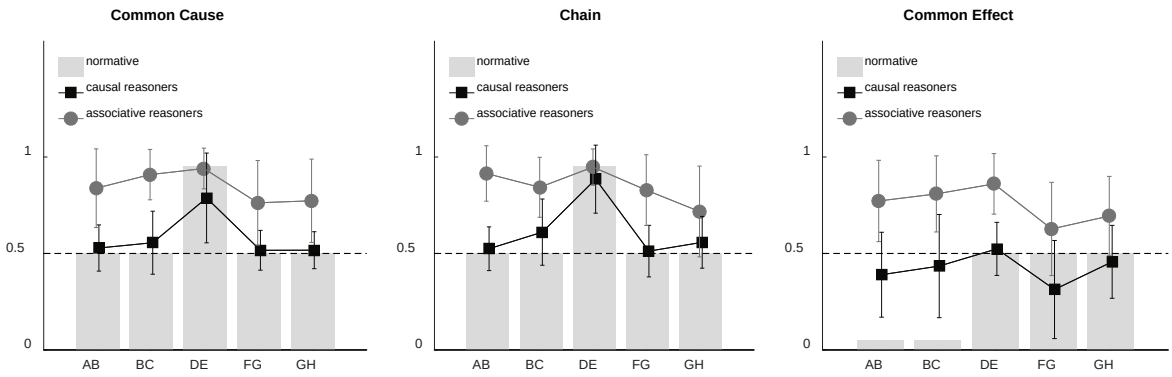


Fig. 2. Mean choice proportions for comparative judgments (AB represents the probability of selecting A vs B). The gray bars show the normative (rational) predictions derived from classical Bayes net principles including the causal Markov condition, and incorporating discounting into the inference for the common effect structure, as reported in Rehder (2014). The line plots show the actual observed aggregated choice responses for participants classified as causal and associative reasoners. The error bars represent a length of two standard deviations, centered on the mean (all adapted from Rehder, 2014).

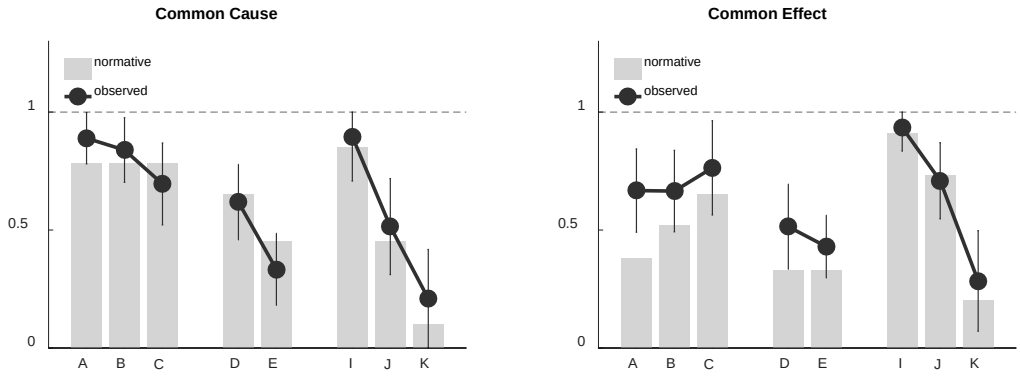


Fig. 3. Mean ratings for absolute judgments. The gray bars show the normative (rational) predictions derived from classical Bayes net principles, as reported in Rehder & Waldmann (2016). The line plots show the mean observed absolute probability judgments. The error bars represent a length of two standard deviations, centered on the mean (all adapted from Rehder & Waldmann, 2016).

2.2. Absolute judgments (dataset 2, Rehder & Waldmann, 2016):

Task

In Rehder and Waldmann (2016), participants were taught either the common cause or the common effect network, similar to Rehder (2014). This study used the same materials, but instead of asking for comparison between scenarios, participants provided the absolute probability (between 0 and 1) of a target variable taking a particular value, for a set of 8 different

situations. The first 5 situations A to E are identical to the previous experiment (see Table 1). The remaining 3 situations (I to K), were different. In these 3 situations, inference needed to be made on the cause in the common cause and the effect in the common effect network, represented by the target variable Z in both cases. Each network type was taught to 48 participants.

Results

Figure 3 summarizes the results. These results correspond to the description-only condition (as opposed to learning from experience) reported in Rehder & Waldmann (2016). For the common cause network (left panel in figure 3), the normative responses reflect the principle of independence for situations A, B, and C when the value of the common cause is known. The moderate downward trend of A-B-C in the observed judgments reflect violations of the principle of independence. The downward trend of D-E and I-J-K in the common cause network reflect normative non-independence between the two effects given that the value of the cause variable is unknown or to be inferred. For the common effect network, the normative increasing slope for A-B-C reflects the effect of 'explaining away', in the presence of alternate causes. The normative equivalence of D-E reflects the independence of the two causes in the absence of any knowledge about the effect. The normative downward trend of I-J-K reflects non-independence while inferring the value of the effect. The deviations between normative and observed patterns, such as the downward trend in A-B-C in the common cause network, the indifference between A-B in the common effect network and the difference in D-E in the common effect network have typically been explained using heuristic explanations.

3. Quantum Probability Model

3.1. General specification

The overarching objective of having a QP model for causal reasoning is as a formalism that will enable the recovery of both normative (CGM) and non-normative (predominantly associative, but also consistent with the other non-normative strategies discussed in Rehder, 2014) influences in participants' performance in the two datasets.

The starting point of a QP model is an assumption of incompatibility, in the present case concerning the mental representations for the three binary variables X, Y, Z, which correspond to a causal reasoning situation (Trueblood & Pothos, 2014; Trueblood et al., 2017). If the X, Y variables are incompatible, then the joint event X&Y does not exist and cannot be assigned a probability. Instead, we have to evaluate the sequential probabilities for X & then Y (i.e., p(X & then Y). This is an appropriate definition for a conjunction for incompatible questions, because it decomposes to the product of a marginal and conditional probability, just like the classical case). The sequential processing of incompatible variables can naturally give rise to order effects in quantum models.

QP theory can be considered a geometric approach to probability where events are defined as subspaces within a vector space (technically a Hilbert space). If we consider the simplest possible representation for incompatible questions, as rays (one dimensional subspaces), then incompatibility means that the rays for the two questions are not orthogonal. We make the additional assumption that the rays corresponding to all the questions in one of Rehder's (2014) causal reasoning scenarios are coplanar. As noted, this is the simplest possible QP causal reasoning model (all variables incompatible, corresponding to rays, and coplanar). It is unlikely

that this 'most quantum' model will be supported in general, but its use here is the most direct test of the hypothesis that quantum principles are relevant in causal reasoning, at least in some cases (see also Trueblood & Pothos, 2014; Trueblood et al., 2017), and also the model's relative simplicity facilitates the individual differences comparison. We illustrate the mechanics of the model through Figure 4. The numbered operations in the figure are referred to using square brackets (e.g. [1]).

The two dimensions for each basis ($\{x_1, x_0\}$, $\{y_1, y_0\}$, $\{z_1, z_0\}$) represent the two values for each binary variable X, Y, and Z (see Figure 4a). Since the causal structures are specified in a single sense (that is, only one value affects the system causally), the values are encoded such that the subscript 1 always indicates the value that is causally linked (e.g. if low interest rates cause high deficits, low interest rates and high trade deficits are encoded as x_1 and y_1 respectively; high interest rates and low trade deficits, which do not influence or experience causal influence, are encoded as x_0 and y_0). One of the variables (in this case, Y) is represented by the standard basis for the 2-dimensional real space (i.e., orthonormal vectors pointing in the direction of the axes of the Cartesian coordinate system), and the basis vectors for X and Z are determined by rotating the standard basis by θ_X and θ_Z respectively. Mathematically, the three bases associated with the three variables are related by rotation matrices R_x for variable X and R_z for variable Z, so that the corresponding basis sets are $\{R_x\, y_1, R_x\, y_0\}$ for X and $\{R_z\, y_1, R_z\, y_0\}$ for Z. These vectors and matrices are given by:

$$y_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (1)$$

$$y_0 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad (2)$$

$$R_x = \begin{bmatrix} cos(\theta_X) & -sin(\theta_X) \\ sin(\theta_X) & cos(\theta_X) \end{bmatrix} \quad (3)$$

$$R_z = \begin{bmatrix} cos(\theta_Z) & -sin(\theta_Z) \\ sin(\theta_Z) & cos(\theta_Z) \end{bmatrix} \quad (4)$$

The degree of rotation between the different subspaces determines the conditional and conjunctive probability relationships between the corresponding variables, that is $p(y_i \mid x_i)$, $p(x_i \mid y_i)$, $p(y_i,x_i)$ and $p(x_i,y_i)$ are all dependent on and can be calculated using θ_X. The three bases are thus related by rotations characterized by the two parameters θ_X and θ_Z.

In QP theory, the state of the system is represented by a state vector ψ**.** For the empirical situations of interest, this state vector represents the mental state of an individual prior to engaging with a causal reasoning problem. The state vector ψ has unit length to maintain probabilities between 0 and 1 (the circles in Figure 4 are unit circles). The probability of a certain variable taking a value (e.g. $p(y_1)$) can be obtained by projecting the state vector (see black dotted line [1] in Figure 4b) onto the basis vector of interest and taking the squared value of the length of the resulting projection (see black bar [2] in Figure 4b). So, the angle between the state vector and a basis vector determines an individual's belief in that variable.

In mathematical terms, the probability of $p(y_1)$ is given by Born's rule:

$$p(y_1) = \| M_{y1}\, \psi \|^2 \quad (5)$$

where M_{y1} is the projection matrix that projects the state vector ψ unto the y_1 subspace (in our case, this is the y_1 basis vector). In this case, projection matrices are given by e.g. $M_{y1} = y_1\, y_1^T$

Conjunctive probabilities (e.g., $p(y_1\&x_1)$ in Figure 4c) are assessed by making successive projections from the belief vector to x_1 [1] and then to y_1 [2]. The final probability $p(y_1\&x_1)$ is then calculated by taking the squared length of the final projection (thick black bar [3]). Note that in this case, the first projection was made onto x_1 and subsequently onto y_1. However, the conjunctive probabilities can also be calculated in the reverse order, that is, by first projecting

onto y_1 and then onto x_1. This would have resulted in a different probability calculation. Unlike classical probability, this model thus differentiates between the operations $p(y_1\&x_1)$ and $p(x_1\&y_1)$, depending on the order of processing the variables. Psychologically, this approach thus predicts order effects in information processing.

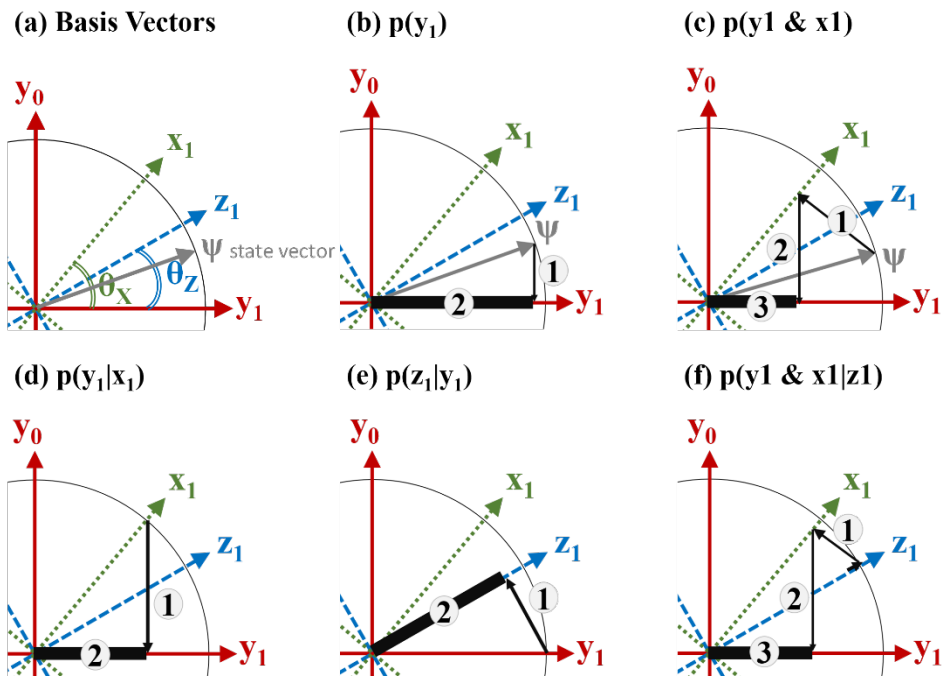


Fig. 4. Details of the QP model. Figure (a) shows the three sets of basis vectors $\{x_0,x_1\}$, $\{y_0,y_1\}$, and $\{z_0,z_1\}$. The three bases are related by rotations characterized by the parameters θ_X and θ_Z. It also shows the state vector (ψ), which is fixed in our modelling (at 45°relative to the horizontal); (b) shows the probability calculation of a single variable; (c) shows conjunctive probability calculations; (d) and (e) show conditional probability calculations; (f) shows conjunctive-conditional probability calculations. The circled numbers show the sequence of operations in each case. All operations start either from the state vector, or in case of conditional probabilities, from the basis vector of the conditional variable. The thick black bar represents amplitude, and the probability is obtained by squaring the amplitude. The circle shown is a unit circle.

To calculate conditional probabilities (e.g. $p(y_1|x_1)$ in Figure 4d), we first assume that the mental state ψ is set to x_1, that is, the mental state is set to the (assumed) known information x_1. Then, we project from this new belief state x_1 onto the basis vector y_1 (operation [1]), and the resulting squared amplitude (operation [2]) gives the final conditional probability $p(y_1|x_1)$. Figure

4e shows the similar conditional calculations for $p(z_1|y_1)$. Finally, Figure 4f shows the calculations for $p(y_1\&x_1|z_1)$. Here we start by assuming that the new belief state is z_1, then project onto the first conjunctive variable (x_1, [1]), and then onto the second (y_1, [2]). The squared length of this final projection [3] gives the required probability. Unlike for CP theory, changing the order of the conjunctive operation (i.e., $p(x_1\&y_1|z_1)$) would again change the result[2].

A basic aspect of all QP models is that a smaller angle of rotation between vectors results in a larger conditional probability, for example, $p(y_1|z_1) > p(y_1|x_1)$ if $\theta_Z < \theta_X$. Also, a general QP prediction is that the order of conjunctive processing matters, that is, $p(x_1 \& y_1) \neq p(y_1 \& x_1)$, unless in the special case where $x_1 \& y_1$ are equidistant from the current belief state. In addition, this simple, 'most quantum' model makes two specific predictions, that do not necessarily hold for more general QP models.

First, when probabilities are conditioned on more than one variable, the order of processing of these given variables is important, since the calculations in the two-dimensional model make all but the last conditional variable redundant. This is the memoryless property. A consequence of this property is order (recency) effects of information processing, for instance, $p(y_1|x_1,z_0) = p(y_1|z_0)$ and $p(y_1|z_0,x_1) = p(y_1|x_1)$ when x_1 is processed first in the former and last in the latter. Note that $p(y_1|x_1,z_0)$ indicates that x_1 is processed first and then z_0. Thus the probabilities $p(y_1|x_1,z_0)$ and $p(y_1|z_0,x_1)$, which would be identical in classical probability, can be different in the two-dimensional QP model.

[2] Examples of some of these calculations are provided in the online supplementary material A

Second, in the two dimensional coplanar QP model $p(y_1|z_1) = p(z_1|y_1)$. This property is known as reciprocity and can be considered an expression of associative thinking in a causal reasoning problem. Both the memoryless property and reciprocity represent predictions of new effects in causal reasoning, from the two dimensional, coplanar QP model, but are not valid in more general QP models, e.g., if variables are represented with higher dimensionality subspaces or positive valued operator measures (POVMs) are employed instead of projections (POVMs capture the situation where measurements are subject to error, and this is covered in greater detail in the discussion section; see Trueblood et al., 2017). It is clearly the case that these two properties are not general properties of causal inference, but are there some circumstances where human behavior reflects the memoryless and reciprocity properties? If yes, this would further inform our understanding of non-normative behavior in causal reasoning. To foreshadow our results, we show that in the simple causal structures examined in this paper, the QP models with these properties provide a better overall representation of human behavior, and especially so in participants that show any form of deviation from the normative CGM approach. That is, we find some evidence of the memoryless property and reciprocity in human causal reasoning under the paradigms examined in this paper.

3.2. Specification of the QP model for the current datasets

In the QP model, the judgments of individuals depend on individual level rotation parameters θ_X and θ_Z, and the projection ordering parameters. We will show that this unified account provides better overall fits, to individual response patterns than the collection of classical models proposed by Rehder (2014). Thus, what appears to be qualitatively different behavior (classified as different heuristic strategies) under the classical framework, may in fact be thought

of as parametric variations of a single QP approach. For both datasets, the model first determines the absolute probability computation for each individual situation (A-H for the first set, and A-E, I-K for the second). Table 2 lists the conditional and conjunctive probability calculations used to infer the absolute probabilities for each of these situations. We distinguish two types of inference situations depending on the number of known variables. The basic principle remains consistent – that all variables, known and unknown, are evaluated, known variables are evaluated before unknown variables, and when more than one known variable is present, individual differences may exist in the order of processing these variables. Finally, when more than one variable is unknown, the target variable, that is the one about which participants are asked to make an inference, is the one processed last.

Absolute judgments with a single known variable:

In situations B, D, E and G, inference on Y is made with only one of the other two variables (either X or Z) being known and the other being unknown. Here there is flexibility regarding how exactly to compute the probability of y_1, depending on whether it is assumed that the participant completely ignores the variable that is unknown or not. We therefore suggest that participants compute the probability of y_1 as $p(y_1 \,\&\, \text{Unknown} = 1 \mid \text{Known}) + p(y_1 \,\&\, \text{Unknown} = 0 \mid \text{Known})$[3]. Recall that conjunctions in QP theory are sequential, and order matters. The specific computations for each of the corresponding situations are detailed in Table 2. The probabilities can be calculated for a given set of values of the rotation parameters for X and Z (relative to the standard basis Y). Individual differences between participants can arise due to

[3] This is just the law of total probability in CP, or QP with the additive interference term set to 0.

differences in the rotation parameters across individuals. The cognitive interpretation of the rotation parameters and the underlying individual differences are elaborated in later sections.

Table 2. Probability calculation under the different scenarios specified in the 2-dimensional QP model. Individual differences in probability estimates can be captured by different rotation parameters, as well as the differences in projection orders in situations with 2 known variables.

Situation	X	Z	Y	# Known Variables	Probability Specification	Individual Differences
A	1	1	target	2	$p(y_1 \mid x_1, z_1)$ or $p(y_1 \mid z_1, x_1)$	Projection order; Rotation parameters
B	?	1	target	1	$p(y_1 \,\&\, x_1 \mid z_1) + p(y_1 \,\&\, x_0 \mid z_1)$	Rotation parameters
C	0	1	target	2	$p(y_1 \mid x_0, z_1)$ or $p(y_1 \mid z_1, x_0)$	Projection order; Rotation parameters
D	1	?	target	1	$p(y_1 \,\&\, z_1 \mid x_1) + p(y_1 \,\&\, z_0 \mid x_1)$	Rotation parameters
E	0	?	target	1	$p(y_1 \,\&\, z_1 \mid x_0) + p(y_1 \,\&\, z_0 \mid x_0)$	Rotation parameters
F	1	0	target	2	$p(y_1 \mid x_1, z_0)$ or $p(y_1 \mid z_0, x_1)$	Projection order; Rotation parameters
G	?	0	target	1	$p(y_1 \,\&\, x_1 \mid z_0) + p(y_1 \,\&\, x_0 \mid z_0)$	Rotation parameters
H	0	0	target	2	$p(y_1 \mid x_0, z_0)$ or $p(y_1 \mid z_0, x_0)$	Projection order; Rotation parameters
I	1	target	1	2	$p(z_1 \mid x_1, y_1)$ or $p(z_1 \mid y_1, x_1)$	Projection order; Rotation parameters
J_X	1	target	0	2	$p(z_1 \mid x_1, y_0)$ or $p(z_1 \mid y_0, x_1)$	Projection order; Rotation parameters
J_Y	0	target	1	2	$p(z_1 \mid y_1, x_0)$ or $p(z_1 \mid x_0, y_1)$	Projection order; Rotation parameters
J		target		2	$J_0 = [p(z_1 \mid y_1, x_0) + p(z_1 \mid x_1, y_0)] / 2$ *or* $J_1 = [p(z_1 \mid x_0, y_1) + p(z_1 \mid y_0, x_1)] / 2$	Projection order; Rotation parameters
K	0	target	0	2	$p(z_1 \mid x_0, y_0)$ or $p(z_1 \mid y_0, x_0)$	Projection order; Rotation parameters

NB. Later on, we use a compact notation for situations e.g. $A(X_1Z_1)$.

Absolute judgments with two known variables:

In situations A, C, F, and H, inference on Y is made conditional on the values of both X and Z. In this case, there is no reason to expect that information about X is processed before or after information about Z, in evaluating conjunctions, so processing order is a free parameter (i.e., a binary switch that governs the order of processing). Table 2 shows two possible calculations for this situation, each defining a different order of processing these known variables. This allows the model to infer the most likely order representation for each participant.

So, for these situations, individual differences can arise both from differences in the rotation parameters and differences in the order of processing the known variables. Similarly, for situations I, J, and K, information about X and Y may be processed in either order to make inferences about Z, and can be a free parameter. The two situations J_X and J_Y are not necessarily symmetric under QP (they may be symmetric if the angle between X-Z and Y-Z is the same), and hence the calculations are performed separately. The average judgment for situation J reported in Rehder & Waldmann (2016) is matched to the average of J_X and J_Y (see Table 2 for details) The projection order for situation J depends on whether the variables with cause (effect) present or absent are considered.

Comparative judgments:

The calculations above yield the desired absolute probabilities for the different situations. This is sufficient for the second data set which captures absolute judgments of probability. For the first dataset, it is necessary to compare probabilities for two situations, denoted S_1 and S_2, where S_1 and $S_2 \in$ {A, B, C, D, E, F, G, H}. In such cases the probabilities $p(y_1 \mid S_1)$ and $p(y_1 \mid S_2)$ are calculated separately, as in Table 2. The final choice proportions between the two are computed based on a softmax decision rule (commonly used to model choice, Daw, O'doherty, Dayan, Seymour, & Dolan, 2006), also utilized by Rehder (2014), so that the probability for selecting S_1 versus S_2 is given by

$$p(S_1 \; vs \; S_2) = \frac{e^{logit(p(y_1 \mid S_1))}}{\sum_{s=S_1,\, S_2}\left(e^{logit(p(y_1 \mid s))}\right)} \qquad (6)$$

This choice proportion is calculated for each of the five problem pairs {AB, BC, DE, FG, GH} covered in experiment 1.

3.3. Hierarchical Bayesian implementation of the naïve (no clusters) QP model (QPN)

We implement the QP model described above using Markov chain Monte Carlo (MCMC) sampling in JAGS[4]. The validity of the inferred parameters is assessed using the $\hat{R}$ statistic (Gelman & Rubin, 1992), which measures the between-chain to within-chain variance. Figure 5 shows the graphical model implementation of the basic QP model. The model requires inference on the rotation parameters (θ_X and θ_Z), and the projection orders for the situations where there are two known variables (i.e., situations A, C, F and H) for each participant. We propose a hierarchical Bayesian model, which allows us to account for individual differences systematically. The first QP model we implement is a naïve model that assumes no clustering of participants.

Parameters θ_Z and θ_X represent the rotation of the Z and X bases from the standard Y basis. Since participants are taught positive causal relationships, the angle between bases of causal parent-child relationships are restricted so that the probability of an effect in the absence of a cause is not judged more likely than the probability of an effect in the presence of the same cause. This restriction is thus placed on θ_Z and on the difference $abs(\theta_X - \theta_Z)$, since these represent the direct parent-child relationships between Z and Y, and between Z and X, in all three network structures. Thus, when $\theta_Z < 45°$, $p(y1|z1) > p(y0|z1)$ and $p(y1|z1) > p(y1|z0)$. Parameters for all participants are drawn from a single hierarchical distribution in the naïve model. Subscript 'i' refers to individuals. For situations A, C, F and H there are two possible projection orders (see Table 2). The priors in the baseline model are uninformative, and the model places an equal prior weight on each projection order for each situation.

[4] Code and documentation for a basic implementation of the QP model is included in the online supplementary material B.

The set of priors used is summarized below, using JAGS notation.

$\mu_Z \sim N(0,1)$ $\mu_D \sim N(0,1)$

$\sigma_Z \sim$ Uniform (0.001,4) $\sigma_D \sim$ Uniform (0.001,4)

$\theta_{Z,i} \sim N(\mu_Z, \sigma_Z)\, T(0, 45°)$ $\theta_{D,i} \sim N(\mu_D, \sigma_D)\, T(-\theta_{Z,i}, 45°)$

$\theta_{X,i} = \theta_{Z,i} + \theta_{D,i}$

The parameters μ_Z and μ_D are the group-level means for the rotation of the Z basis and the difference (D) in the rotation of the Z and X basis, respectively. The parameters σ_Z and σ_D are the group-level standard deviations.

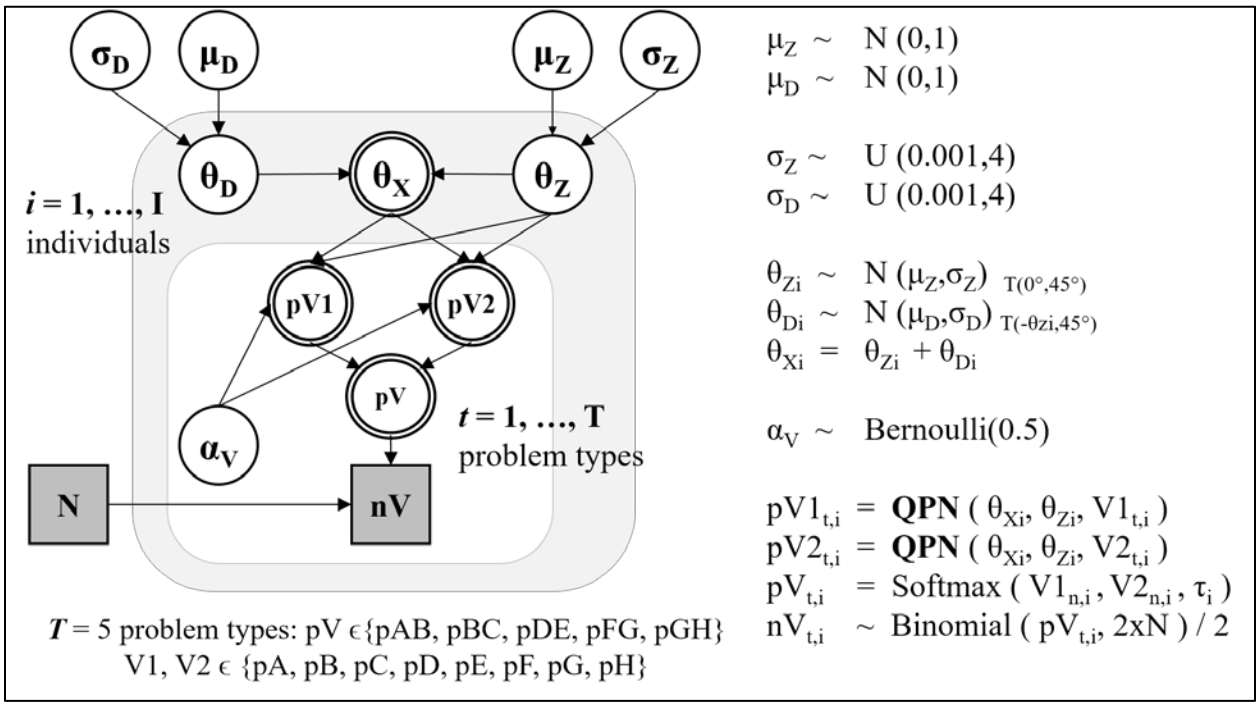


Fig. 5. Expanded graphical model for the naïve QP model (QPN). Circular nodes represent continuous variables, square nodes represent discrete variables, the connecting structures represent the dependencies between variables, and the plates represent repetitions over individuals and problem types. Observed variables are indicated by shaded nodes and unobserved variables by unshaded ones. Finally, probabilistic nodes are indicated by a single border and deterministic nodes with a double border. The situations A to H are represented by V, and the two situations being compared for a question are indicated by V1 and V2. The projection order for any situation V is given by α_V, which takes on a binary value.

Overall, the priors and the structure of the model are identical for all three network structures, CC, CH, and CE. The priors and models used are also identical for both datasets, except for the softmax rule to calculate relative choice proportions, which is required for the first dataset. Thus, differences in inference arising from differences in causal structure are captured completely by the inferred values for the rotation and projection order parameters.

To formulate the specific causal networks investigated, we set the basis for Y (all inference by participants is made on the variable Y) as the standard basis, and two free parameters denote rotations for the basis vectors of X and Z in the 2-dimensional space. Rotations are restricted to the first quadrant, to reduce identifiability issues (for example, a rotation from y_1 of 30° and 330° would result in an identical projection onto y_1). As noted, the location of the state vector is assumed fixed at a neutral position of 45° to the standard basis (y_0-y_1). But, since all inferences are conditional on at least one of the two variables X and Z, the position of the state vector becomes redundant for the presently relevant probability calculations (because of the memoryless property of the model).[5]

The rotation parameters linking the X and Z variables with the Y variable are constrained to be same for all situations A-H in the first dataset and A-E, I-K in the second. This implies that an individual has the same causal representation under all situations, which means that X, Y, Z are incompatible relative to each other in the same way, in different situations. But, we assume that processing order can vary between situations. This can be intuitively understood via an example – let X have the same known value in two situations which differ only in terms of

[5] In situations where the initial state vector has an influence on final probabilities, the location of the state vector could be treated as a free parameter and the QP approach cannot be restricted to the simplest possible one.

whether Z has a causal influence (value 1) or is unknown. The value of Z, and whether this value is in fact known or not, can be envisaged to play a role in whether an individual first processes X or Z.

3.4. Hierarchical Bayesian implementation of the latent clustered QP model (QPC)

We implemented a latent clustering of participants within the QP model, by leveraging the natural structural differences in the projection orders that arise in the QP model. We augmented the naïve QP model with a latent mixture parameter that provides a classification of individuals into multiple clusters. Our latent classification strategy is based on the following assumptions, that can in theory be applied to any causal network.

(1) Each cluster has its own hierarchical parent distribution for all parameters, thus allowing for systematic differences in combinations of rotation and projection order parameters. The priors on the hierarchical distribution for the rotation parameters are identical and relatively uninformed for all groups. Thus, no strong a priori assumptions are made.

(2) Clusters are identified by combinations of projection orders in situations where more than one variable is known, at least one of these variables is known to influence or be influenced in a causal relationship with the variable to be inferred, and the two variables are not interchangeable. The key assumption here is that the order of processing variables plays an important role in individual differences. The discrete nature of possible projection order combinations provides a natural way to segregate behavior.

(3) By implementing this clustering within a Bayesian inference framework, we can obtain the likelihood of each participant belonging to each cluster, implemented by classifying each participant into the modal cluster identified for that participant.

A latent classification parameter (γ) is used to build a mixture model that classifies everyone into one of the clusters[6]. For the first dataset (comparative judgments), there are 3 situations, A, C, and F, each with two possible projection orders, that qualify as unique combinations of projection orders. This results in 8 possible clusters. For the second dataset (absolute judgments), there are 3 situations, A, C, and J, each with two possible projection orders, also leading to 8 possible clusters. Since the clusters are dependent on the projection orders, the optimal number of clusters are automatically chosen by the Bayesian inference mechanism. Foreshadowing results, for the first dataset, the modeling approach results in participants inferred to be distributed across all 8 possible clusters, although a majority (76, 78 and 72 of 105 participants in CC, CH and CE structures respectively) of the participants are inferred to be in 3 of these 8 clusters. For the second dataset, the modeling results indicate participants to be distributed across 7 out of the 8 maximum possible clusters in the CC and CE structures respectively, although a majority (37 and 42 of 48 in CC and CE structures respectively) of the participants are inferred to be in only 3 of these clusters.

4. Specification of heuristic and weighted mixture models based on Rehder (2014)

Along with the QP model described in the previous section, we also implemented the models described in Rehder (2014) within a hierarchical Bayesian framework. Rehder (2014) reported that all four strategies (the normative CGM and the non-normative CONJ, DISAB, and ASSC) *together* could account for participants' behavior. As the present focus is individual differences, we are interested in the extent to which the principles embodied in each strategy are uniformly

[6] Details in online supplementary material C

represented across participants or whether it is the case that different participants focus on different strategies. We thus implement each of the four strategies as separate models within a Bayesian hierarchical framework, as well as a combined weighted mixture model (WTM) that assumes a mixture of the four strategies for everyone (i.e. within-participant mixture)[7]. For the WTM model, the final probability is calculated as a weighted mixture of the 4 strategies, with the weights as free parameters at an individual level. For example, if pFG_{WTM} denotes the relative probability of selecting F (e.g. 0.75 implies selecting F rather than G 75% of the time) under the weighted model, this is calculated as:

$$pFG_{WTM} = w_{CGM}\, pFG_{CGM} + w_{CONJ}\, pFG_{CONJ} + w_{ASSC}\, pFG_{ASSC} + w_{DISAB}\, pFG_{DISAB} \quad (9)$$

The same choice rule for comparing situations is used in all of the models.

5. Applying the models to comparative judgments (dataset 1)

The comparative mean choice proportions for five types of problems aggregated over 4 trials for each type of question i.e. (20 questions per participant) x 105 participants for each of the three network structures x 3 different network structures (common cause, chain, common effect) were used to fit each of the seven models (CGM, CONJ, ASSC, DISAB, WTM, QPN, and QPC) and generate posterior parameter and posterior predictive distributions.

5.1. Model comparison

Figure 6 shows a comparison of the actual behavior and mean posterior predictive values based on the WTM, naïve QP, and clustered QP models. The comparison is separated based on

[7] The detailed specification of these 4 heuristic models is included in the online supplementary material D.

the grouping of causal and non-causal reasoners reported in Rehder (2014). The relatively superior fits of the QP models for the non-causal reasoners can be seen, across all three network structures. The performance of the individual clusters identified by the QPC model is reviewed in greater detail in subsequent sections.

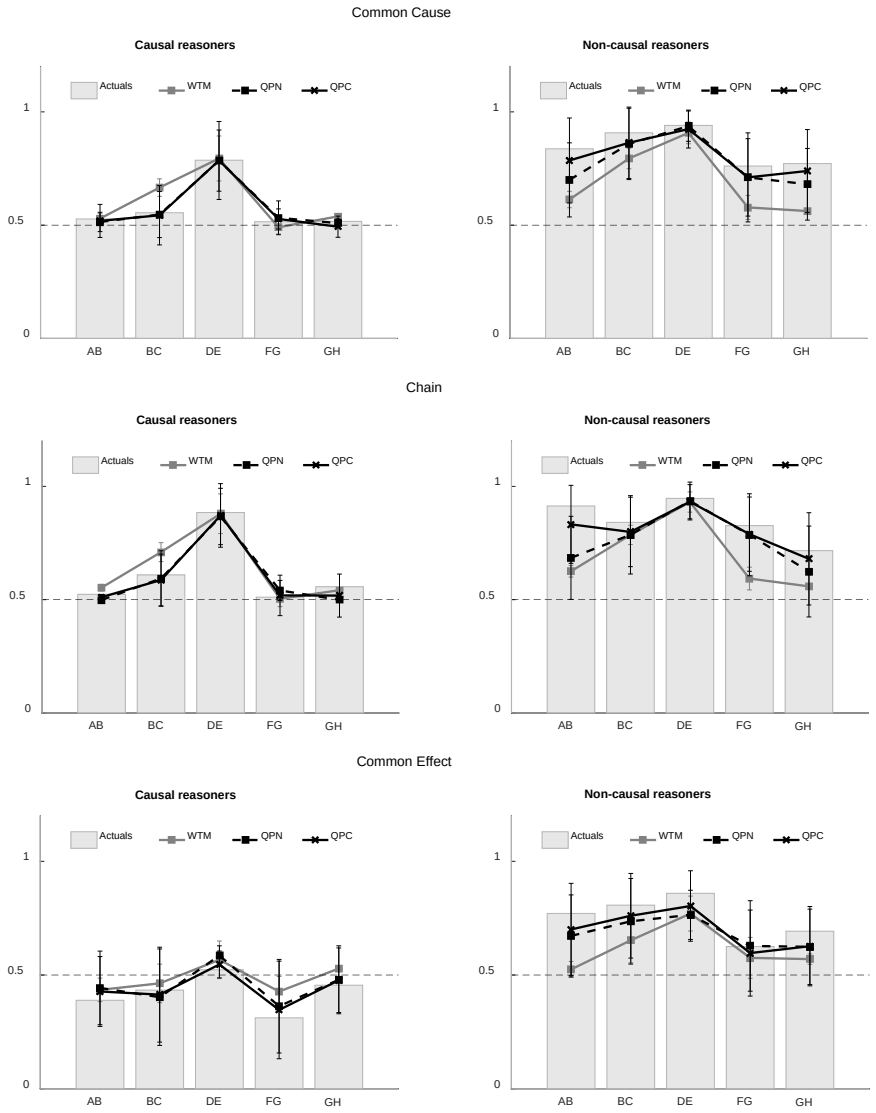


Fig. 6. Comparative judgments (dataset 1): Mean posterior predictive based on the naïve QP (QPN), clustered QP (QPC), and WTM models compared to actual behavior. The gray bars represent the mean actual behavior of the causal and non-causal participants as identified in Rehder (2014).

Model fit is compared using deviance information criteria (DIC), which is a hierarchical modeling generalization of the AIC and BIC, and considers both model fit and complexity for comparing models [8]. A lower DIC value is considered better. DIC is computed as the sum of DBAR (fit of the model to the data) and PD (effective number of parameters of the model). Model fit is also assessed by comparing the correlation, bias, and RMSE of the posterior predictive vs actual observations (Table 3). Table 3 summarizes performance of the six models across the three structures. The clustered QPC model shows the lowest (best) DIC measure across all comparisons, with the slightly higher penalty for complexity (reflected in the higher PD value, which measures the effective number of parameters) being more than compensated for by the superior fit (reflected in the significantly lower DBAR values). The QPC model shows the highest correlation and lowest RMSE of the posterior predictive vs the actual data across all network structures. It also has lower bias (average error between observed data and expected predictions of the model) than most of the models across all networks[9]. Note, the superiority of QPC relative to QPN supports the hypothesis of individual differences in causal reasoning performance, as Rehder (2014) originally suggested.

The individual classical probability or heuristic based models struggle to demonstrate a good fit across all the three network structures, especially in the common effect structure. The CGM, ASSC and DISAB models can show an adequate fit at the aggregate level for the causal reasoners, but not for the non-causal reasoners[10]. The effectiveness of these models is improved using the weighted (WTM) model[11]. Even the WTM model however cannot match the

[8] DIC was calculated using the JAGS DIC package.
[9] See online supplementary material E for a split of the correlation, bias and RMSE by type of reasoner
[10] See online supplementary material F for details of the model fits segregated by type of reasoner
[11] See online supplementary material G for details of the inferred strategy weights for the WTM model

performance of the QP models for all three structures. Interestingly, the clustered QPC model shows lower model complexity, measured by the effective number of parameters (PD), than the naïve QPN model, and at a comparable level to the WTM model.

Table 3. Model Comparison. DIC is the deviance information criteria and measures the fit (DBAR) and complexity (PD; measures the effective number of parameters) of the models, DIC = PD + DBAR. Lower values of DIC, DBAR and PD are desirable. The QP model has the lowest values for all 3 network structures. The correlation, bias and RMSE are measured between the mean of the posterior predictive values generated by the model and the actual data. Higher correlation, lower bias and RMSE are desirable. Numbers in bold indicate the best model per that criterion.

			Common Cause			
Model	DIC	PD	DBAR	Correlation	Bias	RMSE
CGM	1975	**53**	1922	0.57	-0.074	0.192
CONJ	2957	58	2899	0.35	-0.103	0.281
ASS	1796	77	1719	0.71	**-0.007**	0.153
DISAB	1809	114	1695	0.73	-0.025	0.150
WTM	1830	89	1741	0.72	-0.015	0.152
QP (naïve)	1595	111	1484	0.92	-0.017	0.091
QP (clustered)	**1432**	88	**1344**	**0.94**	-0.014	**0.077**
			Chain			
Model	DIC	PD	DBAR	Correlation	Bias	RMSE
CGM	2024	**47**	1977	0.58	-0.095	0.207
CONJ	2523	72	2451	0.38	-0.036	0.242
ASS	1772	79	1692	0.70	-0.023	0.159
DISAB	1866	76	1790	0.69	**0.012**	0.165
WTM	1800	88	1712	0.71	-0.022	0.157
QP (naïve)	1608	131	1477	0.86	-0.034	0.118
QP (clustered)	**1383**	70	**1313**	**0.93**	-0.022	**0.087**
			Common Effect			
Model	DIC	PD	DBAR	Correlation	Bias	RMSE
CGM	2501	87	2415	0.43	-0.042	0.246
CONJ	2551	77	2474	0.49	**-0.003**	0.251
ASS	2443	**69**	2374	0.59	0.069	0.242
DISAB	2456	82	2374	0.46	-0.021	0.238
WTM	2159	82	2078	0.78	0.017	0.194
QP (naïve)	1725	109	1616	**0.92**	0.01	0.112
QP (clustered)	**1627**	90	**1537**	**0.92**	**0.003**	**0.109**

5.2. Individual level comparison of QP and WTM models using Bayes Factors

The data was used to fit the QPC and WTM model within a single Bayesian inference framework, which allows us to calculate the pairwise Bayes Factors using the product space method (Lodewyckx, Kim, Lee, Tuerlinckx, Kuppens, & Wagenmakers, 2011). A Bayes Factor comparison (ratio of likelihoods of the observed data given each model) considers model fit and complexity and measures the relative evidence for one model over another. A Bayes factor (BF) of one indicates equal evidence for both models under consideration. Only BFs that are larger than 10 (or smaller than 10^{-1}) are considered strong, and those that are larger than 100 (or smaller than 100^{-1}) are considered decisive (Jeffreys, 1961). We measure the pairwise Bayes factor[12] at an individual level for each participant. Figure 7 shows the individual level log Bayes factors (LBF) in favor of the QPC model versus WTM. 73 of the 315 participants had strong (LBF > log(10)) or decisive evidence (LBF > log(100)) in favor of the QPC model, and 33 in favor of the WTM model.

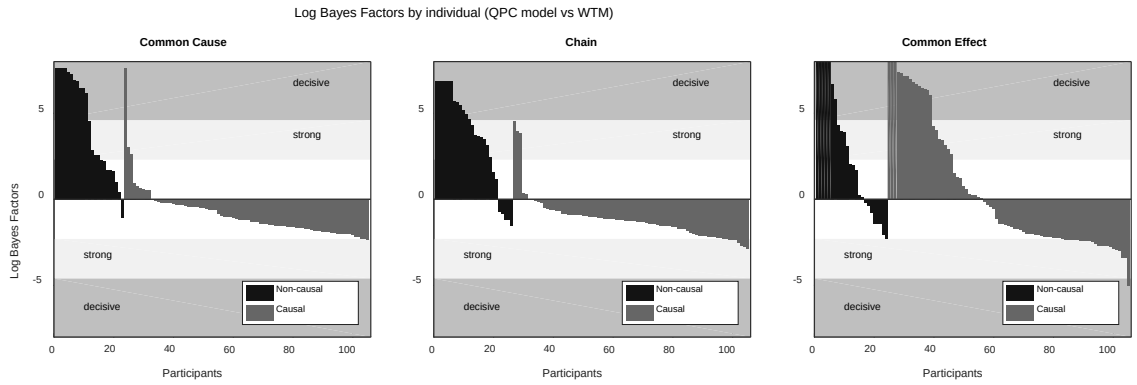


Fig. 7. Individual differences model comparison: Pairwise log Bayes factors (LBF) by individual. Black bars represent non-causal reasoners and gray bars causal reasoners (based on Rehder's, 2014, classification). Log Bayes factors are plotted in favor of the QPC model vs WTM model, so that positive values are in favor of QPC and negative values are in favor of WTM. The plots show regions of *strong* (LBF > log(10) in favor of either model) and *decisive* (LBF > log(100) in favor of either model) evidence based on LBF.

[12] Details of the methodology and pseudopriors used are provided in the online supplement K.

We examined whether there are data patterns that the QPC model can account for but the WTM model cannot (and vice versa), using the 73 and 33 participants for which the LBF strongly favors one of the two models. Figure 8 highlights the mean posterior predictive values from the two models compared to the actuals, for these participants. Points along the diagonal reflect most accurate model behavior. The top panel shows that the WTM model sometimes finds it hard to account for judgments which the QPC model describes reasonably well for the 73 participants. The bottom panel shows that for the 33 participants where the LBF strongly favors the WTM model, the WTM model does not generally provide superior fits compared to the QPC model.

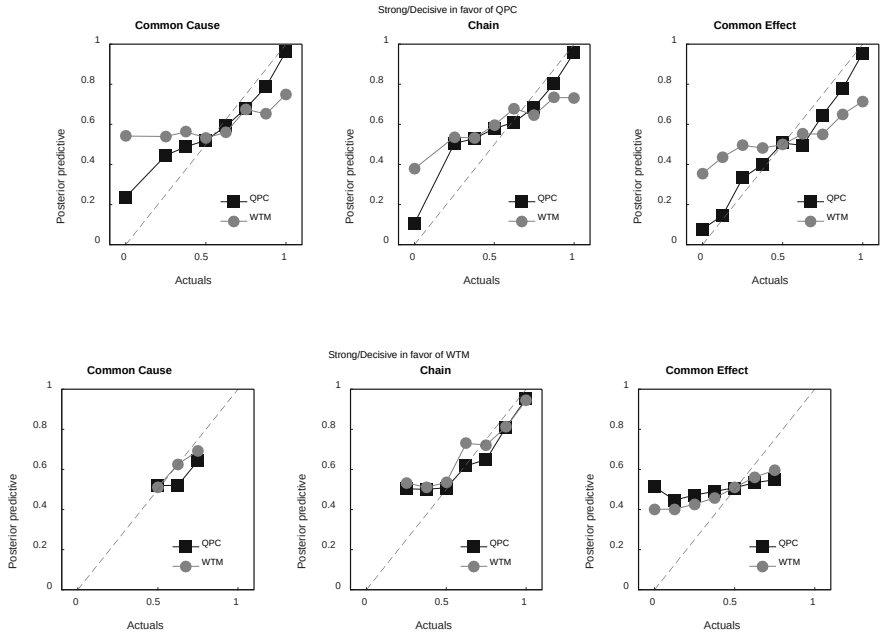


Fig. 8. Mean posterior predictive vs actuals for participants where $BF_{QP\ vs\ WTM} > 10$ (top panel, 73 participants) and $BF_{WTM\ vs\ QP} > 10$ (bottom panel, 33 participants). This figure shows all question types combined (comparisons AB, BC, DE, FG, and GH) for the three network structures. The model posterior predictive values are similar in the bottom panel. On the other hand, the top panel shows that the WTM model struggles to fit behavior for participants where the QPC model shows superior BF. This is especially true for extreme values of comparative judgments that the QPC model can account for, but the WTM model does not.

5.3. Clustering for comparative judgments

The clustered QPC model groups participants probabilistically into clusters based on the preferred projection orders. The clustering emerges as a natural property of the QPC model, with the only prior attribution being differences in projection orders. By examining the resulting clusters and inferred rotation parameters for each, we can attribute interpretable behavioral characterizations to each cluster. By grouping participants based on the modal cluster, we identify multiple behavioral patterns that show nuanced differences compared to a broad classification into just causal versus non-causal reasoners. Table 4 show the latent QPC based clustering of participants for each of the three networks. The rows show the 8 clusters and the behavioral characterization for each. Participants in each cluster may show some variation of rotation parameters within the cluster, however the variation between clusters is significantly higher than within cluster. The columns show the mean value of the X and Z rotation parameters in each cluster. The rotations do not reveal the direction of causality but the strength of the bidirectional association between the variables, with a lower rotation implying higher associative strength. A rotation of less than 45° indicates a positive association, and more than 45° indicates a negative association between the target variable Y and the relevant variable X or Z (independence between 2 variables is characterized by an angle of 45°). The rotation parameters along with the distinct behavioral patterns and projection orders highlight how the different clusters plausibly reflect distinct behaviors[13]. The remaining columns show the number of participants out of 105 in each network that are classified into each cluster. The columns C^R and N^R show participants classified as causal and non-causal respectively by Rehder (2014).

[13] See online supplement H for the joint posterior density of the rotation parameters for each cluster.

Table 4. Clustering under the QPC model. Note that since clustering takes place within a probabilistic inference framework, the grouping is reported based on the cluster that was the modal cluster for each participant.

Common Cause network			**Mean inferred rotation (degrees)** [4]		**Latent Classification** [5, 6]		
Cluster [1,2]	Conflict (focused variable[3])	Symmetry	X	Z	C^R	N^R	total
1. A_{Z1}-C_{Z1}-F_{Z0}	Immediate	No	0.3	24.4	31	1	32
2. A_{Z1}-C_{Z1}-F_{X1}	Present	No	44.1	42.4	12	-	12
3. A_{Z1}-C_{X0}-F_{Z0}	Absent	No	9.6	27.5	6	2	8
4. A_{Z1}-C_{X0}-F_{X1}	Distant	No	31.0	10.6	1	5	6
5. A_{X1}-C_{Z1}-F_{Z0}	Immediate	Yes	16.1	17.4	31	1	32
6. A_{X1}-C_{Z1}-F_{X1}	Present	Yes	37.8	38.5	-	2	2
7. A_{X1}-C_{X0}-F_{Z0}	Absent	Yes	4.8	16.3	-	7	7
8. A_{X1}-C_{X0}-F_{X1}	Distant	Yes	31.5	21.6	1	5	6
Chain network							
Cluster	Conflict	Compression	X	Z	C^R	N^R	total
1. A_{Z1}-C_{Z1}-F_{Z0}	Immediate	No	0.3	17.9	24	-	24
2. A_{Z1}-C_{Z1}-F_{X1}	Present	No	41.5	40.2	7	-	7
3. A_{Z1}-C_{X0}-F_{Z0}	Absent	No	31.7	30.8	12	1	13
4. A_{Z1}-C_{X0}-F_{X1}	Distant	No	24.8	1.5	-	9	9
5. A_{X1}-C_{Z1}-F_{Z0}	Immediate	Yes	10.2	11.5	33	8	41
6. A_{X1}-C_{Z1}-F_{X1}	Present	Yes	25.7	31.0	-	2	2
7. A_{X1}-C_{X0}-F_{Z0}	Absent	Yes	29.3	17.6	3	1	4
8. A_{X1}-C_{X0}-F_{X1}	Distant	Yes	25.3	13.4	-	5	5
Common Effect network							
Cluster	Conflict	Discounting	X	Z	C^R	N^R	total
1. A_{Z1}-C_{Z1}-F_{Z0}	Immediate	Ignore alternate cause	42.7	43.8	41	3	44
2. A_{Z1}-C_{Z1}-F_{X1}	Distant	Ignore alternate cause (monotonicity violated)	63.7	43.5	8	-	8
3. A_{Z1}-C_{X0}-F_{Z0}	Immediate	Anti-discounting (monotonicity violated)	18.8	24.9	4	8	12
4. A_{Z1}-C_{X0}-F_{X1}	Distant	Anti-discounting	40.1	29.0	-	7	7
5. A_{X1}-C_{Z1}-F_{Z0}	Immediate	Anti-discounting (weak positive product synergy)	36.7	36.3	3	5	8
6. A_{X1}-C_{Z1}-F_{X1}	Distant	Discounting (strong negative product synergy)	61.2	39.0	7	-	7
7. A_{X1}-C_{X0}-F_{Z0}	Immediate	Anti-explaining away (weak positive product synergy)	38.8	34.7	2	1	3
8. A_{X1}-C_{X0}-F_{X1}	Distant	Explaining away (strong negative product synergy)	68.4	44.0	16	-	16

Notes:

1. Clustering is based on the preferred projection orders, i.e. whether final projection is on Z or X
2. Projection order: for each situation, we show where the final projection is made *from*, e.g. for A_{Z1} the final projection is made from Z1 for situation A. Recall, for Rehder's (2014) dataset, inference is always made on variable Y. For all networks, *immediately* related variables are ones that are in a direct child or parent relationship (e.g., in the CC network Z-X and Z-Y), otherwise they are *distant*. For all three networks Z is the immediately related variable and X the distant variable. For the CC network the common cause is Z and the effects X, Y. For the CH network the effect is

Y and X, Z are causes. For the CE network, the common effect is Z. The target cause is the one about which reasoners have to make an inference (Y), the alternate cause (X) is the other one.

3. The focused variable is the one from which the last projection is made.
4. Mean rotation parameters (degrees) are the inferred means for the clusters, but note that there exist individual differences within clusters as well (see online supplement H for details of the cluster level distributions)
5. Latent classification shows the number of participants classified based on the modal cluster
6. C^R (causal) and N^R (non-causal) participants as identified by post hoc clustering in Rehder (2014)

Behavior in each cluster (a latent group identified by the QP model) can be characterized by considering what happens in each situation (i.e., a particular combination of X, Z, e.g., $C(Z_1X_0)$). An empirical effect relevant for all networks is *conflict*. Conflict situations are those where a causal relationship appears to be violated, that is, an effect is absent even in the presence of its cause, or an effect is present even when none of its causes are present. Plausible heuristics that people may use to respond to such situations include focusing on the variable that is *present* (i.e., the projection order for situation S would be e.g. S_{X1} or S_{Z1}, so that the last projection is from X_1 or Z_1), on the variable that is *absent* (i.e., the projection order for situation S would be e.g. S_{X0} or S_{Z0}, so that the last projection is from X_0 or Z_0), on the variable that is *immediately* related (Z; note this is normative for CC and CH networks), or on the variable that is *distant* (X).

For clusters in the common cause network, situations $C(Z_1X_0)$ and $F(Z_0X_1)$ represent conflict. Another relevant effect is *symmetry*, in situation $A(Z_1X_1)$, whereby reasoners make inferences on the effect Y, based on an assumption of association/ equivalence between the effects sharing the cause (symmetry is a subset of non-independence). For example, in situation A, if the cause X has value 1, then by symmetry, cause Y is also 1.

Regarding the chain network, $C(Z_1X_0)$ and $F(Z_0X_1)$ are conflict situations. Situation $A(Z_1X_1)$ potentially reflects a *compression* effect, whereby inference at one end of the chain is based on the other end of the chain, ignoring intermediate nodes. For example, in situation A, if the cause X has value 1, then by compression, effect Y is also 1.

In the common effect network, only situation $F(Z_0X_1)$ represents conflict. In situations $A(Z_1X_1)$ and $C(Z_1X_0)$ the common effect (Z) is present and there is a question of the association between the X, Y causes. *Positive* (the two causes enhance each other), *negative* (the two causes undermine each other), *zero*, or *ambiguous* (it is not known to the reasoner whether the causes enhance or undermine each other) *product synergy* can exist between causes (Wellman & Henrion, 1993; Druzdzel & Henrion, 1993). Product synergy captures the sign of conditional dependence between the two causes. Negative synergy is related to the effects of discounting and explaining away, whereas positive synergy is related to anti-discounting. Explaining away is defined as when the probability of the target cause being present is strictly reduced, as the alternate cause goes from being absent to ambiguous to present (Rehder & Waldmann, 2016), and is normative. Explaining away is a strong form of discounting. Negative product synergy may not necessarily result in the conditions for explaining away, but still lead to discounting, specifically active discounting (no distinction between the alternate cause being absent or ambiguous, but lower probability of target cause when the alternate cause is present) and passive discounting (no distinction between the alternate cause being present or ambiguous, but higher probability of target cause when the alternate cause is absent).

Detecting conflict and symmetry effects depends only on projection order parameters, but for explaining away or types of discounting in the CE network, one also has to consider QP rotation parameters. For inferences based on the last projection being from the alternate cause X, $\theta_X > 45°$ reflects explaining away, based on presence or absence of this alternate cause. $\theta_X < 45°$ indicates that reasoners correlate alternate causes, so this corresponds to anti-explaining away. For inferences based on the last projection being from common effect Z, this corresponds to ignoring the alternate cause altogether. A further distinction is relevant in situations A and C in

the CE network, for which active discounting (or anti-discounting) occurs when the alternate cause is considered in situation A but not in C, and passive discounting (or anti-discounting) occurs when the alternate cause is considered in situation C but not in A. Here it is both θ_Z and $(\theta_X - \theta_Z)$ that determine whether we have discounting or anti-discounting. In some such cases, there may be only partial evidence for such effects.

For the CE network, also relevant is the possibility of ambiguous product synergy, where the nature of relationship between causes is unknown (Wellman & Henrion, 1993). If the value of the alternate cause is unknown, reasoners may have difficulty thinking of how the two causes are related. So, we may have scenarios where the probability of the target cause is judged to be especially high or low when the alternate cause is unknown (situation B) compared to when it is known (situations A and C), violating the monotonicity of probability of target cause conditional on the probability of alternate cause. Specifically, we could have probability of A vs B > 0.5 and probability of B vs C < 0.5 (i.e. probability of B is low), or probability of A vs B < 0.5 and probability of B vs C > 0.5 (i.e. probability of B is high). This behavior indicates discounting in one comparison and anti-discounting in the other. Note, it would be difficult to explain such behavior classically, but in the QP model, when the value of the alternate cause (X) is unknown, the probability calculations for situation B (see Table 2) make it possible to violate monotonicity.

So far we focused on situations A, C, F, since these guide the clusters. Additionally, for all three networks, we can interpret judgments in situations D, E, where Z is unknown, using θ_Z. When θ_Z is small, the judged relative probability of D vs E is very sensitive to θ_X, whereas as θ_Z approaches 45°, situations D and E are judged as almost equally likely, regardless of θ_X. Intuitively, clusters with high values for θ_Z imply that individuals treat X and Y as relatively independent.

Armed with the precise characterization of each cluster from the QP model parameters (projection order and rotation angles), we can look at the clusters identified by the Bayesian latent classification and explore individual differences. We start with normative behavior. Aspects of normative behavior (such as independence) are reflected in most clusters, but purely normative clusters are relatively rare. For both the CC and CH networks, only cluster 1 is normative. Normative behavior in the CC network requires no symmetry and resolution of conflict situations using the immediate variable (as in cluster 1), and in the CH structure requires no compression and resolution of conflict situations under the immediate variable (again as in cluster 1). These conditions essentially imply that the final projections should not be from the distant variable X when both variables are known.

In the CE network, normative behavior depends on both projection order and rotation parameters, and in fact, none of the clusters are completely normative across the requirements from both the Markov condition and discounting. A completely normative cluster in the CE network would have projection orders as in cluster 7 (A_{X1}-C_{X0}-F_{Z0}), required for the Markov condition, and rotation parameters as in cluster 8, $\theta_X > 45^0$ for discounting or explaining away. Projection orders different from cluster 7 violate the Markov condition by ignoring the alternate cause when the effect is present or by considering the alternate cause when the effect is absent. On the other hand, $\theta_X < 45^0$ implies a positive synergy between the alternate and target cause, and thus results in behavior that is opposite to discounting or explaining away.

One way to characterize the degree of normativity of a cluster is to calculate the difference (RMSE) between the mean proportions for participants in the cluster, and the predictions from the normative CGM network (as reported in Rehder, 2014). These RMSE values are calculated for each cluster, and also summarized across groups of clusters. These

groups are based on how participants in a cluster handle conflict, as well as symmetry (in CC), compression (in CH), and discounting (in CE). These grouped RMSE values are shown in table 5[14]. Higher RMSE indicates greater deviation from normative behavior.

Table 5. RMSE shows the difference between mean proportions for participants in each cluster (based on modal cluster) and the normative predictions of the CGM model (Rehder, 2014). Lower values of RMSE closer to zero indicate normative-like behavior, and higher values indicate greater deviation from normative behavior. The RMSE values are based on a comparison of all 5 comparative probability judgments for each network structure. The clusters are grouped based on conflict and symmetry behavior for common cause, conflict and compression behavior for chain, and conflict and discounting behavior for common effect structures. Value in bold indicate large differences between groups.

Common Cause		**Clusters**	**RMSE versus Normative**	**#Participants**
Conflict	Immediate	1,5	**0.13**	64
	Present	2,6	**0.24**	14
	Absent	3,7	**0.32**	15
	Distant	4,8	**0.37**	12
Symmetry	No	1,2,3,4	0.22	58
	Yes	5,6,7,8	0.22	47
Chain		**Clusters**	**RMSE versus Normative**	**#Participants**
Conflict	Immediate	1,5	**0.14**	65
	Present	2,6	**0.22**	9
	Absent	3,7	**0.27**	17
	Distant	4,8	**0.38**	14
Compression	No	1,2,3,4	0.22	53
	Yes	5,6,7,8	0.22	52
Common Effect		**Clusters**	**RMSE versus Normative**	**#Participants**
Conflict	Immediate	1,3,5,7	0.38	67
	Distant	2,4,6,8	0.39	38
Discounting	Ignore alternate cause	1,2	**0.32**	52
	Anti-discounting	3,4,5,7	**0.53**	30
	Discounting	6,8	**0.30**	23

For the CC network, conflict seems to play a large role in influencing normative behavior, with an RMSE of 0.13 for immediate, 0.24 for present, 0.32 for absent, and 0.37 for

[14] See online supplement I for RMSE by individual cluster, to see how the different clusters represent different aspects of normative versus non-normative behavior.

distant resolutions of conflict situations. Presence (mean RMSE 0.22) or absence (mean RMSE 0.22) of symmetry does not seem to be a significant factor regarding variability in normative behavior. The CH network shows similar trends, with an RMSE of 0.14 for immediate, 0.22 for present, 0.27 for absent, and 0.38 for distant resolutions of conflict situations. Presence (mean RMSE 0.22) or absence (mean RMSE 0.22) of compression similarly does not seem to be a significant factor. For the CE network, conflict does not appear to make normative behavior more or less likely. The mean RMSE is similar for clusters where discounting is demonstrated (mean 0.30) and where the alternate cause is ignored (mean 0.32), but much higher (i.e. greater deviation from normative) when anti-discounting behavior is demonstrated, with a mean RMSE of 0.53, as is expected.

We now proceed to consider individual differences more generally. For the CC network, most participants were assigned to clusters 1 & 5. For both clusters, final projection is from the immediate variable Z, not the distant variable X, under both conflict situations C and F. Behaviorally, this suggests that reasoners are taking into account exclusively or primarily just the immediate variable, which is normative, since the causes are independent. Indeed, 64 out of 82 causal reasoners from Rehder (2014) are clustered in clusters 1 & 5. However, 12 out of 82 of Rehder's (2014) causal reasoning are in cluster 2, where the final projection is from the present (=1) variable under both conflict situations C, F, which is normative only for situation C but not F. In the latter case, this possibly indicates matching behavior (Evans & Lynch, 1973) or some other attentional bias. Rehder's (2014) non-causal reasoners are 10/23 in clusters 4 & 8 and 9/23 in clusters 3 & 7. The latent clustering indicates a diversity of ways in which non-normative behavior can arise, mostly in terms of which variable is focused on and whether symmetry is adhered to (non-normative) or not. For example, for conflict situations C, F, there is evidence for

a bias to focus both on the distant variable (clusters 4 & 8) and the absent one (clusters 3 & 7). Overall, of the 105 participants, 47 were clustered in groups that showed symmetry under situation A, whereas the remaining 58 did not.

For the CH network, the latent clustering again reveals a large number of participants in clusters 1 & 5, for which the final projection is from the immediate variable Z, not the distant one X, in both conflict situations C, F (which is normative, because of independence). Of Rehder's (2014) causal reasoners, 57/79 are in clusters 1 & 5. The non-normative participants are again distributed in several smaller clusters, showing a variety of behavioral patterns. For example, clusters 4 & 8, representing 14/26 of Rehder's (2014) non-causal reasoners, reflect last projection from the distant variable in conflict situations C, F, violating the Markov condition. There was a large percentage of compression instances (52/105).

For the CE network, as above, final projection from the immediate variable Z, rather than the distant one X, is the normative behavior in conflict situation F and 50/81 out of Rehder's (2014) causal reasoners are in corresponding clusters. However, equally 17/24 of non-causal reasoners were also included in such clusters. Unlike for the CC and CH networks, together with the conflict situation we have to consider (normative) discounting behavior (determined from the rotation parameters) before a response pattern is considered as normative. So, in clusters 3, 4, 5, and 7 there is evidence for anti-discounting and account for 21/24 of non-causal reasoners from Rehder (2014). In clusters 6, 8 parameter analysis reveals mostly discounting behavior and 23/81 causal reasoners from Rehder (2014) are in these clusters. However, 49/81 of Rehder's (2014) causal reasoners were also in clusters 1, 2, which show no discounting (the alternate cause is basically ignored, when its presence should inform the inference from the target cause). No discounting is not normative, so this is a situation where the present approach deviates from

Rehder's (2014) classification in a substantial way. Analogously, 41 of Rehder's (2014) causal participants were assigned in cluster 1, but again in that cluster there is clear evidence for non-normative behavior (the alternate cause is ignored, so no discounting).

Interestingly, θ_Z values are relatively higher for the CE network (mean 37°) compared to the CC (mean 24°) and CH (mean 17°), implying that the primitive probabilistic association between cause and effect is affected by the network structure. These values imply that the average conditional probability of an effect given the cause, independent of other variables, would be approximately 0.65 in the CE, but closer to 0.9 in the CH and 0.83 in the CC network. Figure 9 shows the mean actual behavior, and posterior predictive based on the QPC and WTM models for participants in each of the 8 clusters inferred from the QPC model, for the three network structures. The merit of the QPC model is evident in, for instance, groups 4, 6, and 7 (about 14% of participants) in the common cause and chain networks, and groups 2, 4, and 8 (about 30% of participants) in the common effect structure, where the WTM finds it difficult to account for these behavioral data. In most of the remaining clusters, the QPC model is still slightly better, or comparable, to the WTM model. Overall, a merit of the QPC approach is that it enables a precise characterization of behavior in each cluster. First, we can trace the exact way in which violations of normative prescription arise. For example, the link between potential conflict and non-normative behavior is a novel inference from the QP model that is not apparent from the WTM or any of the underlying heuristic models. Second, the QP model distinguishes between ignoring alternate causes, discounting, and anti-discounting, and between active and passive drivers of discounting or anti-discounting. Clusters also reflect differences in responding to uncertainty, attenuating network structures through mechanisms such as compression and symmetry, and primitive strength of probabilistic cause-effect associations.

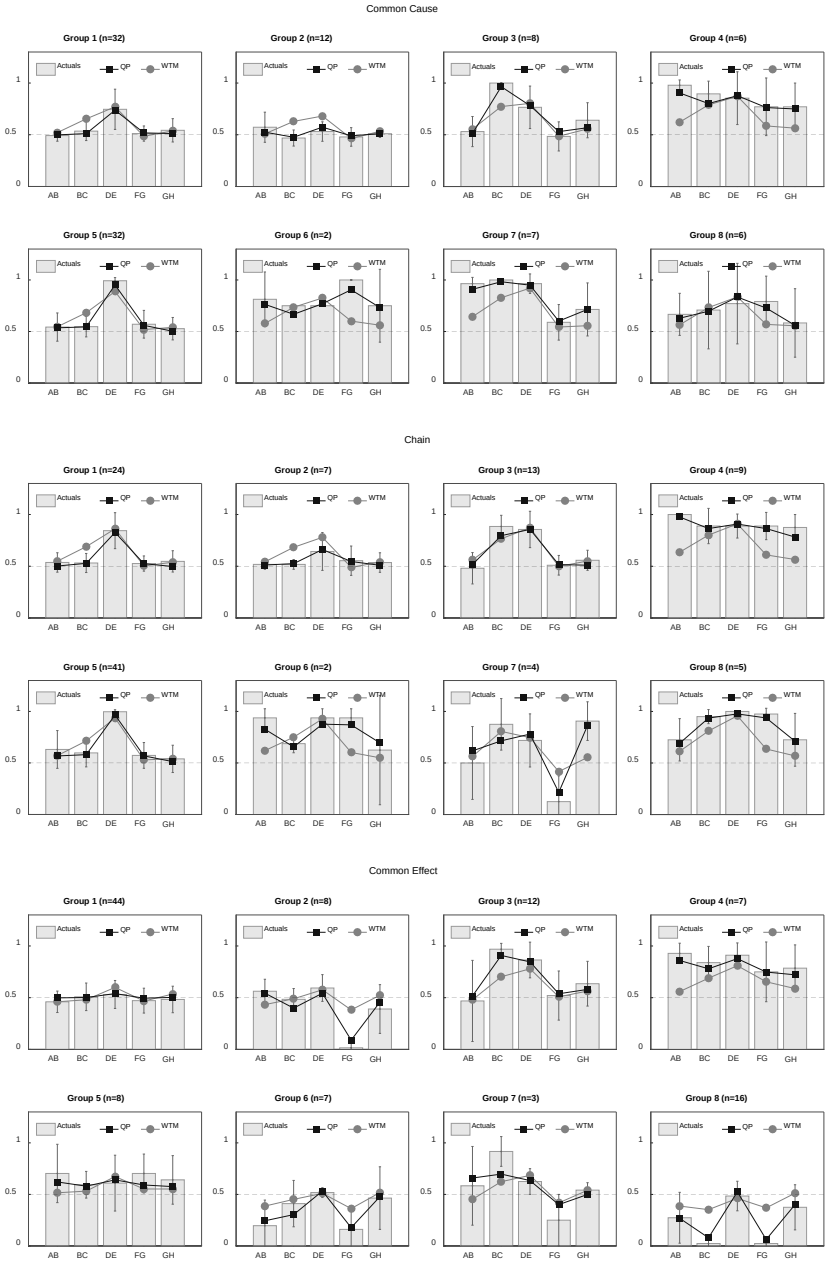


Fig. 9. Comparative judgments (dataset 1): Mean posterior predictive based on the QPC and WTM models compared to the mean actual behavior (gray bars). The data is shown separately for each cluster identified by the QPC model. The plot titles show the number of participants in each cluster (based on modal cluster values).

All these factors are employed to understand heterogeneity amongst participants. This behavioral characterization is conducted using clusters which are inferred without the application of any specific prior knowledge[15].

6. Applying the models to absolute judgments (dataset 2)

6.1. Model comparison

Figure 10 shows the mean actual (from Rehder & Waldmann, 2016) and mean posterior predictive absolute judgments based on the WTM, QPN, and QPC models. Table 6 compares the models using DIC, correlation, bias, and RMSE. The posterior predictive generated by the QPN, QPC, and WTM models are highly similar. The WTM model has better DIC, correlation and error metrics for the common cause network, whereas the QPC model has better DIC, correlation, bias and RMSE metrics for the common effect network. Even though the differences are not as pronounced as in the first dataset, the main question presently is whether the simple 2-dimensional QP model can generate plausible absolute probability judgments in a variety of situations: the QPN model is comparable to the WTM for absolute judgments.

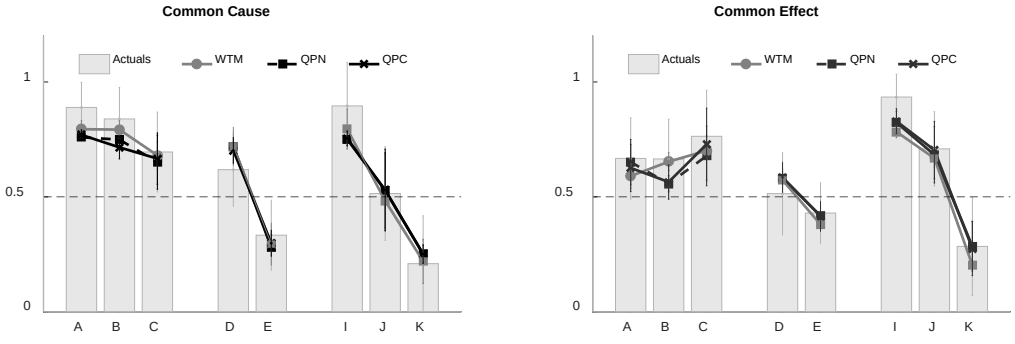


Fig. 10. Absolute judgments (dataset 2): Mean posterior predictive based on QPN, QPC, and WTM models compared to the mean actual behavior (gray bars).

[15] Since the QPC model predictions are based on comparative judgments, a satisfactory plausibility check involves examination of the intermediate absolute probability judgments inferred by the model. This is presented in the online supplement J.

Table 6. Model Comparison. DIC is the deviance information criterion and measures the fit (DBAR) and complexity (PD; measures the effective number of parameters) of the models; DIC = PD + DBAR. Lower values of DIC, DBAR and PD are desirable. The correlation, bias and RMSE are measured between the mean of the posterior predictive values generated by the model and the actual data. Higher correlation, lower bias and RMSE are desirable. Numbers in bold indicate the best model per that criterion.

			Common Cause			
Model	DIC	PD	DBAR	Correlation	Bias	RMSE
WTM	**-417**	104	**-520**	**0.97**	**-0.03**	**0.07**
QPN (naïve)	-260	87	-347	0.95	-0.04	0.09
QPC (clustered)	-319	**49**	-368	0.96	-0.04	0.09
			Common Effect			
Model	DIC	PD	DBAR	Correlation	Bias	RMSE
WTM	-270	68	-339	0.96	-0.05	0.08
QPN (naïve)	-325	108	-432	0.96	-0.04	0.07
QPC (clustered)	**-427**	**63**	**-490**	**0.97**	**-0.03**	**0.06**

Table 7. Clustering under the QPC model. Since clustering takes place within a probabilistic inference framework, the grouping is reported based on the cluster that was the modal cluster for each participant. Latent classification shows the number of participants classified based on the modal cluster. See the Table 4 caption for terminology.

Common Cause network				**Mean inferred rotation (degrees)**		**Latent Classification**
Cluster	Conflict-J	Conflict-C	Symmetry	X	Z	total
1. A_{Z1}-C_{Z1}-J_0	Absent	Immediate	No	14.4	14.0	6
2. A_{Z1}-C_{X0}-J_0	Absent	Distant	No	32.1	16.2	3
3. A_{X1}-C_{Z1}-J_0	Absent	Immediate	Yes	7.4	19.7	11
4. A_{X1}-C_{X0}-J_0	Absent	Distant	Yes	21.2	17.5	1
5. A_{Z1}-C_{Z1}-J_1	Present	Immediate	No	14.3	12.1	9
6. A_{Z1}-C_{X0}-J_1	Present	Distant	No	36.8	17.3	1
7. A_{X1}-C_{Z1}-J_1	Present	Immediate	Yes	5.5	21.5	17
8. A_{X1}-C_{X0}-J_1	Present	Distant	Yes	23.0	20.1	-
Common Effect network				**Mean inferred rotation (degrees)**		**Latent Classification**
Cluster	Conflict-J	Discounting		X	Z	total
1. A_{Z1}-C_{Z1}-J_0	Absent	Ignore alternate cause (monotonicity violation)		47.1	27.4	1
2. A_{Z1}-C_{X0}-J_0	Absent	Anti-discounting		43.1	13.2	1
3. A_{X1}-C_{Z1}-J_0	Absent	Anti-discounting		19.1	34.9	1
4. A_{X1}-C_{X0}-J_0	Absent	Anti-explaining away		31.2	23.0	-
5. A_{Z1}-C_{Z1}-J_1	Present	Ignore alternate cause		22.1	22.0	11
6. A_{Z1}-C_{X0}-J_1	Present	Anti-discounting (monotonicity violation)		44.7	35.8	3
7. A_{X1}-C_{Z1}-J_1	Present	Discounting		41.2	0.3	23
8. A_{X1}-C_{X0}-J_1	Present	Focus on alternate cause (but zero product synergy)		43.8	16.7	8

6.2. Clustering for absolute judgments

Table 7 shows the modal clustering of participants into one of the 8 natural clusters. Figure 11 compares the actual and posterior predictive absolute judgments[16]. An important addition to this dataset is situation J, where one of the effects (causes) in the common cause (common effect) network is 0 and the other is 1, and inference needs to be made on the common cause (common effect). Since both known variables are immediately related to the variable on which inference is being made, the conflict situation J can be resolved by either selecting the variable that is present (1) or absent (0). The projection orders in the QP models thus either select the variable that is 0 (J_0) or 1 (J_1) to make the final projection (Table 2). Since the networks were taught to participants based on the influence of cause present variables, under the QP model, projection order J_1 comes closest to normative behavior.

We rely on the same principles as in section 5.3 to characterize the clusters, in terms of the various effects of conflict, symmetry, discounting, etc.. In the CC network, for conflict situation C, 43/48 participants are clustered under groups where focus is on the variable in an immediate relationship with the target variable (normative). Regarding conflict situation J, evidence for normative behavior is more mixed, with 27 out of 48 participants clustered under groups where focus is on the present (clusters 5-8) vs absent (clusters 1-4) variable. In clusters 5-8 the probability of J is higher than 0.5 and in clusters 1-4 lower than 0.5. Therefore, interestingly, focus on the present vs. absent variable and the corresponding change in normative status appears to impact the perception of probability for the situation.

[16] The online supplement L compares the joint posterior density of the rotation parameters, by cluster.

In the common effect network, 45 out of 48 participants are clustered under groups where conflict in situation J is resolved based on the present (clusters 5-8) vs absent (clusters 1-4) variable. Overall, most of the participants here are in clusters 5, 7, or 8 and, even though the focus on the present cause in J is normative, in some of these clusters there is no evidence of discounting (in situations A, B, C), which is not normative. In cluster 5 participants ignore the alternate cause (i.e., no discounting), demonstrating higher absolute probabilities in situations A, B, and C. Participants in cluster 8 focus on the alternate cause X, but the inferred rotation parameters (θ_X very close to 45°) show that there is almost no product synergy. Hence the presence, absence, or ambiguity of the alternate cause X seem to provide almost no information to participants about Y, resulting in absolute judgments of probability close to 0.5 for Y conditional on X1 or X0, in situations A, B, and C. That is, focus on X does not appear to translate into discounting.

However, in cluster 7, there is mixed evidence for discounting. In situation C, the low value of θ_Z (mean 0.3°) leads to a high inferred probability of the target cause ($p(Y \mid Z_1)$ is very high), i.e. we observe discounting relative to situation B. Under situation A though, inference is made based on a relatively uninformative alternate cause (mean θ_X = 41.2°), so there is no drop in probability in situation A compared to B (i.e., no discounting). The non-normative responses (a small minority) to situation J in clusters 1-4 cannot be effectively captured by the WTM model, but are well characterized by the QPC model.

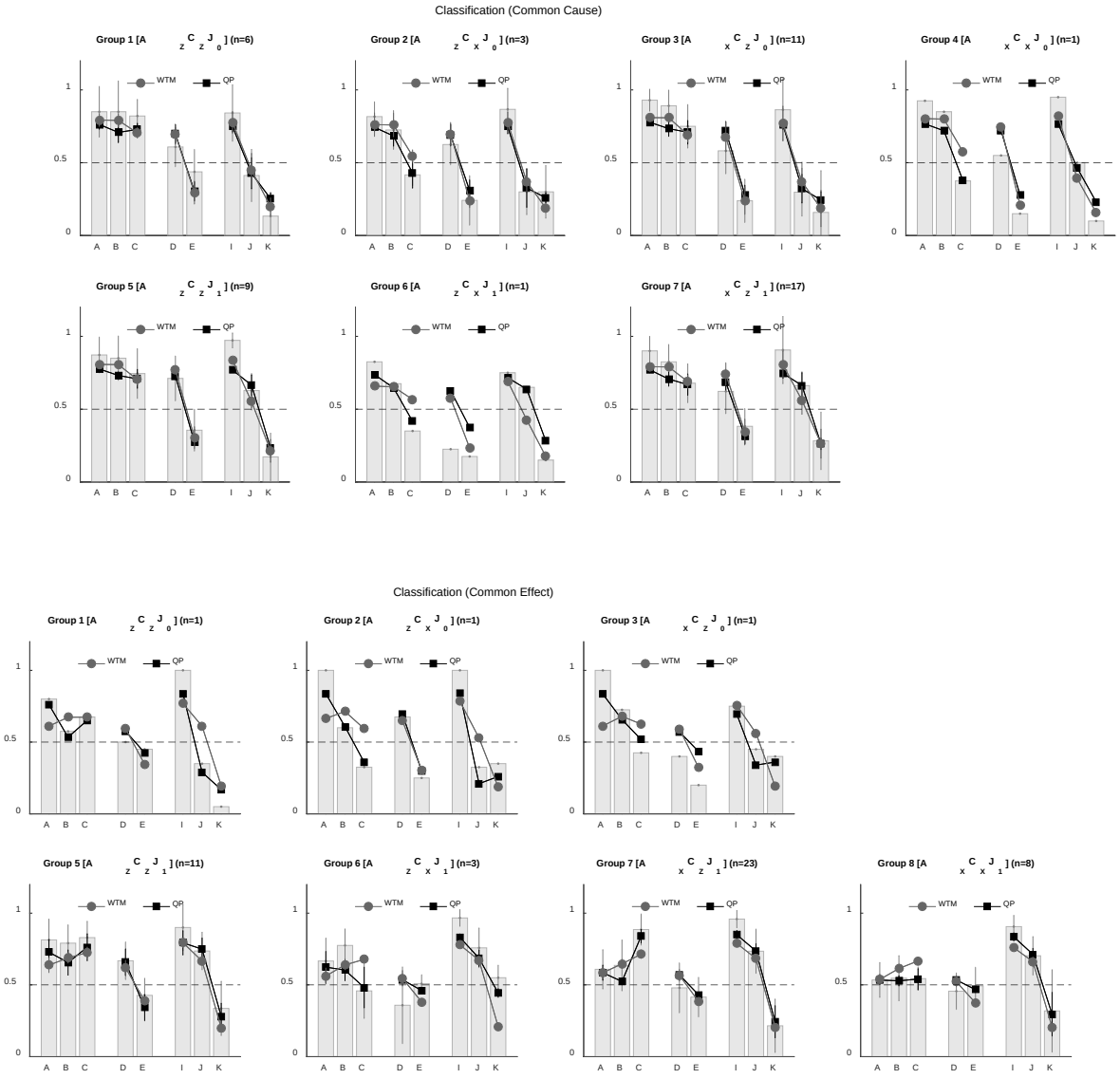


Fig. 11. Absolute judgments (dataset 2): Mean posterior predictive based on the QPC and WTM models compared to the mean actual behavior (gray bars). The data is shown separately for each cluster identified by the QPC model. The plot titles show the number of participants in each cluster (based on modal cluster values).

7. General Discussion

Evidence of non-classicality in causal reasoning has been abundant, but a rigorous formalization of non-classical principles, to complement normative ones, has been elusive. An important step in that direction was Rehder's (2014) proposal of three heuristic strategies,

intended to complement normative principles. Because of a track record of applications of QP into decision situations that conflict with classical normative prescription, we were interested in whether a QP model could be proposed for causal reasoning, that would encompass some or all of the heuristic strategies Rehder (2014) proposed. We outlined the 'most quantum' possible QP model, based on the simplest possible representations (coplanar rays) for the relevant variables, to account for the data in Rehder (2014) and Rehder and Waldmann (2016). Rehder (2014) developed his empirical tests specifically to test violations of classicality so we reasoned that, if such a 'most quantum' QP model were to have a chance, these would be suitable datasets to test it on.

The advantages of this approach are that this simple QP model enables a clear expression of quantum principles potentially applicable in causal reasoning and corresponding performance predictions at the individual participant level not apparent in existing models of causal reasoning. For instance, the QP model allows inferences on mental processing order of causal variables and indeed on which variables are utilized or not. Moreover, regarding individual differences, Rehder (2014) already reports many interesting insights. We aimed for a more detailed analysis, guided by latent classification of the QP parameters. This allowed identification of several determinants of individual differences, including conflict situations in all networks, symmetry in CC networks, compression in CH networks, and a range of discounting, anti-discounting and intermediate behaviors in CE networks.

Overall, this paper provides one of the most detailed individual differences analyses in causal reasoning, as implemented within a hierarchical Bayesian framework (cf. Busemeyer, Wang, & Shiffrin 2015), and involving classical (normative) influences, heuristic models, and the novel framework for non-classical behavior, QP. The analyses encompassed individual

differences in both Rehder's (2014) extensive data set, involving 315 participants, multiple causal structures, and manipulating content, as well as for a second dataset measuring absolute probability judgments (Rehder & Waldmann, 2016), involving 48 participants. For these particular datasets, the QP model could account parsimoniously for behavior. In the following sections, we discuss applicability, scope, strengths and weaknesses of the QP approach to investigating individual differences in causal reasoning.

7.1. Complexity and parsimony of QP models

A concern of QP models is whether they introduce too much complexity. Presently, the same QP model was used for all three network structures and for absolute and comparative judgments of probability, without strong informative priors. Moreover, at an individual level the QP model has two rotation parameters, a latent class parameter, and a projection order parameter. A complete weighted mixture of strategies model (WTM) as proposed by Rehder (2014) involves 17 parameters (4 mixture weights, 3 CGM, 3 CONJ, 2 ASSC, and 5 DISAB parameters).

We have explored the questions of fit and complexity in multiple technical ways. The QP model showed categorically better model fits. Across the three network structures, the posterior predictive of the QPC model shows a correlation with the actual data ranging from 0.92 to 0.94, whereas the classical probability and heuristic models showed correlations ranging from 0.35 to 0.78, for the first dataset using relative probability judgments. A Bayes factor comparison with the WTM model showed that for 66% of the participants, the Bayes factor did not exceed 10 (representing strong evidence) in favor of either model. Of the remaining 34% participants, 23% showed a BF greater than 10 in favor of the QPC model and less than 11% showed a BF greater

than 10 in favor of the WTM model. The Bayes Factor comparison considers model complexity and fit, and demonstrates that any complexity introduced by the QPC is justified by fit quality and posterior predictive generated (cf. Busemeyer, Wang, & Shiffrin, 2015). The DIC comparison between models shows that the complexity of the QPC model is comparable to that of the WTM model, but the QPC model provides a better fit in most cases. In comparison to Rehder's (2014) models, the QPC model does not show higher RMSE due to overfitting, which is typical of overly complex models. Notwithstanding these technical points, the conceptual advantage of QPC is that, compared to a combination of four different strategies, QPC applies a single set of principles to all three network structures, and can account for individual differences across both types of reasoners Rehder (2014) postulated, based on differences in model parameters. Specifically, individual behavior patterns that have been conceptualized as involving qualitatively different causal models can instead be accommodated through alterations of continuous parameters within a single framework. Also, the latent classification emerges from the natural structure of the 2-dimensional QP model.

To show that the QPC model is identifiable, we conducted a parameter recovery exercise. The correlation between simulated and recovered data is 0.96, and between simulated and recovered parameters is 0.94 and 0.93 for the rotation parameters for X and Z respectively. Further details of the parameters and data recovery are included in the online supplement M. We also generated data using the WTM model and attempted to recover this using the QPC model. The recovery of WTM simulated data by QPC is reasonable (correlation 0.81), but not as robust as the fit of human data by the QPC model (correlation 0.93), demonstrating that the improved fit of the QPC model is not because it is significantly more flexible. There are clearly some data

configurations that the WTM model produces that the QPC model finds difficult to recover. Details can be found in supplement N.

7.2. Extensions of the QP model

The QP model explored in this paper was restricted to a two-dimensional model, which is the simplest possible QP model for the relevant causal structures. Such a low dimensionality approach requires that the three variables under consideration (X, Y, and Z) are incompatible and so should be evaluated sequentially. This model and the underlying assumptions are adequate for Rehder's (2014) and Rehder and Waldmann's (2016) datasets, including aspects of normative behavior (Section 3.3). However, it seems clear that individuals can produce normative behavior under a wider range of inference problems, that is, they form compatible (i.e., classical) representations of events in some situations. For example, classical probability models provide good accounts of causal reasoning when individuals learn causal relationships through observation or have access to statistical information (e.g., contingency tables; e.g., Cheng, 1997). Thus, a general theory of causal reasoning must be able to account for both Bayesian and non-Bayesian influences.

In the present paper, we focus on the question of whether there are causal reasoning situations when the simplest, 'most quantum' possible model can provide an adequate account of behavior. An advantage of the present approach is that it provides an arguably purer test to the hypothesis that incompatibility can have an explanatory role in causal reasoning. The two-dimensional model discussed in this paper can be considered 'fully' quantum since all variables are incompatible. For the type of problems explored in this paper, models with more compatible events would either be 4 dimensional, which assumes that two of the three variables may be

jointly evaluated, or 8 dimensional, which assumes all variables can be evaluated jointly. All models can also be elaborated through the application of POVMs, which retain the incompatibility assumption, but represent a situation no longer constrained by reciprocity. However, adopting the simplest possible QP model, as we did, has conceptual and technical advantages regarding the individual differences and model comparison analyses with the heuristic proposals reported in Rehder (2014).

Recently, Trueblood et al. (2017) proposed a general framework for human inference (including causal inference) using QP theory. Their approach presents a hierarchy of mental representations, from 'fully' quantum to 'fully' classical, that could be adopted in different situations. In this hierarchy of models, moving from the lowest level to the highest involves changing assumptions about compatibility (i.e., how joint events are represented). Models with more incompatible events have lower dimensionality than models with more compatible events. Thus, levels in this hierarchy correspond to probabilistic models of different dimensionality with the highest dimensional model being 'fully' classical and the lowest dimensional model being 'fully' quantum. Because transitions within the hierarchy involve changing assumptions about compatibility, there is a transparent, detailed way for how the different models in the hierarchy are linked. In particular, Trueblood et al. (2017) showed that as participants gained familiarity with a causal reasoning task, there was a shift in the best fit model from quantum to classical. These results suggest that when people have more experience in a reasoning task their mental representations become more normative in nature.

7.3. Concluding Comments

We have considered the cases of two datasets, specifically constructed to explore non-classical influences in causal reasoning, where individual differences are apparent and the presence of both causal and associative (or other sub-groups) of reasoners exist. In these two cases, a simple QP model provides a parsimonious and comprehensive account of the different types of behaviors, only with a change in parameter values. By implementing a classical probability model, a quantum probability model, and several heuristic models within a common hierarchical Bayesian framework that allows evaluation of individual differences and exhaustive model comparisons, we were able to provide detailed conclusions regarding the relevance of quantum principles in causal reasoning and a range of effects which can be used to characterize non-normative causal judgments.

Author's Note

The authors thank Bob Rehder for sharing his data and details of the classical models used. PKM and JST were supported by NSF grant SES-1556415. JST was also supported by NSF grant SES-1556325. JV was supported by grants 1230118 and 1534472 from NSF's Methods, Measurements, and Statistics panel and grant 48192 from the John Templeton Foundation. EMP was supported by Leverhulme Trust grant RPG-2015-311 and H2020-MSCA-IF-2015 grant 696331. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the funding agencies.

References

Aerts, D., Broekaert, J., Gabora, L., & Sozzo, S. (2013). Quantum structure and human thought. Behavioral and Brain Sciences, 36(3), 274-276.

Aerts, D., Gabora, L., & Sozzo, S. (2013). Concepts and their dynamics: A quantum-theoretic modeling of human thought. Topics in Cognitive Science, 5(4), 737-772.

Aerts, D., Sozzo, S., & Veloz, T. (2016). New fundamental evidence of non-classical structure in the combination of natural concepts. Phil. Trans. R. Soc. A, 374(2058), 20150095.

Atmanspacher, H., & Filk, T. (2010). A proposed test of temporal nonlocality in bistable perception. Journal of Mathematical Psychology, 54, 314–321.

Brainerd, C., Wang, Z., & Reyna, V. (2013). Superposition of episodic memories: Overdistribution and quantum models. Topics in Cognitive Science, 5(4).

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

Busemeyer, J. R., Wang, Z., Pothos, E. M., & Trueblood, J. S. (2015). The conjunction fallacy, confirmation, and quantum theory: Comment on Tentori, Crupi, and Russo (2013).

Busemeyer, J. R., Wang, Z., & Shiffrin, R. M. (2015). Bayesian model comparison favors quantum over standard decision theory account of dynamic inconsistency. Decision, 2(1), 1.

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

Cowell, R. G., Dawid, A. P., Lauritzen, S. L., & Spiegelhalter, D. J. (1999). Probabilistic networks and expert systems. Statistics for Engineering and Information Science. Springer-Verlag New York Inc.

Daw, N. D., O'doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. Nature, 441(7095), 876.

Druzdzel, M. J., & Henrion, M. (1993). Intercausal reasoning with uninstantiated ancestor nodes. In Proceedings of the Ninth international conference on Uncertainty in artificial intelligence (pp. 317-325). Morgan Kaufmann Publishers Inc..

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2010). Neglect of alternative causes in predictive but not diagnostic reasoning. Psychological Science, 21 (3), 329-336.

Fernbach, P. M., & Rehder, B. (2013). Cognitive shortcuts in causal inference. *Argument & Computation*, *4*(1), 64-88.

Fernbach, P. M., & Sloman, S. A. (2009). Causal learning with local computations. *Journal of experimental psychology: Learning, memory, and cognition*, *35*(3), 678.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. Statistical science, 457-472.

Goodman, N. D., Ullman, T. D., & Tenenbaum, J. B. (2011). Learning a theory of causality. Psychological Review, 118 (1), 110-119.

Griffiths, T. L., & Tenenbaum, J. B. (2009). Theory-based causal induction. Psychological Review, 116 (4), 661-716.

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. Trends in Cognitive Sciences, 14, 357-364.

Hagmayer, Y., & Waldmann, M. R. (2002). A constraint satisfaction model of causal learning and reasoning. In W. D. Gray & C. D. Schunn (Eds.), Proceedings of the twenty-fourth annual conference of the cognitive science society (p. 405-410). Mahwah, NJ: Erlbaum.

Jeffreys, H. (1961). Theory of probability (3rd edition). New York: Oxford University Press.

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

Jones, M. & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. Behavioral and Brain Sciences, 34, 169, 231.

Kemp, C., Goodman, N. D., & Tenenbaum, J. B. (2010). Learning to learn causal models. Cognitive Science, 34 (7), 1185{1243.

Kim, J. H., & Pearl, J. (1983). A computational model for causal and diagnostic reasoning in inference systems. In Proceedings of the 8th international joint conference on artificial intelligence (ijcai) (pp. 190-193).

Koehler, J. J. (1996). The base rate fallacy reconsidered: Descriptive, normative, and methodological challenges. *Behavioral and brain sciences*, *19*(01), 1-17.

Lee, M. D. (2011). How cognitive modeling can benefit from hierarchical Bayesian models. Journal of Mathematical Psychology, 55(1), 1-7.

Lee, M. D. & Webb, M. R. (2005). Modeling individual differences in cognition. Psychonomic Bulletin & Review, 12, 605-621.

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

Lober, K., & Shanks, D. R. (2000). Is causal induction based on causal power? critique of Cheng (1997). Psychological Review, 107 (1), 195-212.

Lodewyckx, T., Kim, W., Lee, M. D., Tuerlinckx, F., Kuppens, P., & Wagenmakers, E. J. (2011). A tutorial on Bayes factor estimation with the product space method. Journal of Mathematical Psychology, 55(5), 331-347

Marr, D. (1982). Vision: a computational investigation into the human representation and processing of visual information. San Francisco: W. H. Freeman.

Mistry, P. K., Trueblood, J. S., Vandekerckhove, J. & Pothos, E. M. (2015). A latent-mixture quantum probability model of causal reasoning within a Bayesian inference framework. Proceedings of the 37th Annual Conference of the Cognitive Science Society. Austin, TX: Cognitive Science Society.

Moreira, C., & Wichert, A. (2014). Interference effects in quantum belief networks. Applied Soft Computing, 25, 64-85.

Oaksford, M. & Chater, N. (2009). Précis of Bayesian rationality: the probabilistic approach to human reasoning. Behavioral and Brain Sciences, 32, 69-120.

Park, J., & Sloman, S. A. (2013). Mechanistic beliefs determine adherence to the Markov property in causal reasoning. Cognitive psychology, 67(4), 186-216.

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

Pearl, J. (2014). The deductive approach to causal inference. Journal of Causal Inference, 2 (2).

Evans, J. St.B. T., & Lynch, J. S. (1973). Matching bias in the selection task. British Journal of Psychology, 64, 391–397.

Pothos, E. M., & Busemeyer, J. R. (2009). A quantum probability model explanation for violations of “rational” decision making. Proceedings of the Royal Society B, 276(1665), 2171–2178.

Pothos, E. M., Busemeyer, J. R., & Trueblood, J. S. (2013). A quantum geometric model of similarity. Psychological Review, 120(3), 679.

Pothos, E. M. & Busemeyer, J. R. (2013). Can quantum probability provide a new direction for cognitive modeling? Behavioral & Brain Sciences, 36, 255-327.

Pothos, E. M., & Trueblood, J. S. (2015). Structured representations in a quantum probability model of similarity. Journal of Mathematical Psychology, 64, 35-43.

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

Rehder, B., & Burnett, R. C. (2005). Feature inference and the causal structure of categories. Cognitive Psychology, 50(3), 264-314.

Rehder, B., & Waldmann, M. R. (2016). Failures of explaining away and screening off in described versus experienced causal learning scenarios. Memory & cognition, 1-16.

Rottman, B. M., & Hastie, R. (2014). Reasoning about causal relationships: Inferences on causal networks. Psychological bulletin, 140(1), 109.

Rottman, B. M., & Hastie, R. (2016). Do people reason rationally about causally related events? Markov violations, weak inferences, and failures of explaining away. *Cognitive psychology*, *87*, 88-134.

Russell, S. J., Norvig, P., Canny, J. F., Malik, J. M., & Edwards, D. D. (2003). *Artificial intelligence: a modern approach* (Vol. 2). Upper Saddle River: Prentice hall.

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.

Sozzo, S. (2015). Conjunction and negation of natural concepts: A quantum-theoretic modeling. Journal of Mathematical Psychology, 66, 83-102.

Stanovich, K. E. & West, R. F. (2000). Individual differences in reasoning: implications for the rationality debate? Behavioral and Brain Sciences 23, 645-726.

Tenenbaum, J. B., & Griffiths, T. L. (2001). The rational basis of representativeness. In *Proceedings of the 23rd annual conference of the Cognitive Science Society* (pp. 1036-41). Lawrence Erlbaum Associates.

Tenenbaum, J. B., Griffiths, T. L., & Kemp, C. (2006). Theory-based Bayesian models of inductive learning and reasoning. Trends in Cognitive Sciences, 10 (7), 309{318.

Trueblood, J., & 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.

Trueblood, J. S., Mistry, P. K., & Pothos, E. M. (2015). A Quantum Bayes Net Approach to Causal Reasoning. Contextuality from Quantum Physics to Psychology, 6, 449.

Trueblood, J. S., & Pothos, E. M. (2014). A Quantum Probability Approach to Human Causal Reasoning. In Proceedings of the 36th annual conference of the cognitive science society (pp. 1616-1621)

Trueblood, J. S., Yearsley, J. M., & Pothos, E. M. (2017). A quantum probability framework for human probabilistic inference. *Journal of Experimental Psychology: General*, *146*(9), 1307.

Tucci, R. R. (1995). Quantum Bayesian nets. International Journal of Modern Physics B, 9(03), 295-337.

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.

Waldmann, M. R., Cheng, P. W., Hagmayer, Y., & Blaisdell, A. P. (2008). Causal learning in rats and humans: A minimal rational model. The probabilistic mind. Prospects for Bayesian cognitive science, 453-484.

Wang, Z., Busemeyer, J. R., Atmanspacher, H., & Pothos, E. M. (2013). The potential of using quantum theory to build models of cognition. Topics in Cognitive Science, 5(4), 672-688.

Wellman, M. P., & Henrion, M. (1993). Explaining 'explaining away'. IEEE Transactions on Pattern Analysis and Machine Intelligence, 15(3), 287-292.

White, P. A. (2005). The power pc theory and causal powers: Comment on Cheng (1997) and Novick and Cheng (2004). Psychological Review, 112 (3), 675-682.

A. Examples of QP calculations

See attached excel sheet VAL.xlsx

B. Code and documentation for a basic implementation of the QP model

See attached txt file QPN.txt

C. Hierarchical Bayesian implementation of the clustered quantum probability (QPC) model

Based on the cluster definition assumptions, reasoning in each of the three networks in the first dataset can be classified into 8 possible clusters, depending on the combination of projection orders for the situations A, C and F. These clusters are shown in Table C1.

The latent classification prior for each individual γ_i is given a uniform categorical prior over these 8 clusters. The clusters are defined top-down (as in a latent class analysis), in terms of the distribution over projection orders. Hence, each individual is classified into the cluster whose projection orders provide the highest likelihood of generating the behavior observed for that individual.

Table C1. 8 possible clusters in the first dataset based on combinations of projection orders for the situations A, C, and F.

	Preferred projection orders		
Cluster	**A**	**C**	**F**
1	A_Z	C_Z	F_Z
2	A_Z	C_Z	F_X
3	A_Z	C_X	F_Z
4	A_Z	C_X	F_X
5	A_X	C_Z	F_Z
6	A_X	C_Z	F_X
7	A_X	C_X	F_Z
8	A_X	C_X	F_X

D. Hierarchical Bayesian specification of CGM, CONJ, DISAB, ASSC models

Rehder (2014) reported that all four strategies (the normative GCM one and the non-normative CONJ, DISAB, and ASSC ones) *together* could account for participants' behavior. As the present focus is individual differences, we are interested in the extent to which the principles embodied in each strategy are uniformly represented across participants or whether it is the case that different participants focus on different strategies. We thus implement each of the four strategies as separate models, as well as a combined weighted mixture model that assumes a mixture of the four strategies for everyone (i.e. within-participants mixture).

Normative Causal Graphical Model [CGM]

Figure D1 shows the normative CGM for the common cause, chain and common effect network structures. The normative CGM defines the joint and conditional probability of the variables as in a causal Bayes net. The model has the following free parameters:

c = prior probability of the primary (independent) cause(s), $0 \leq \boldsymbol{c} \leq 1$

m = strength of the individual causal links from cause to effect, $0 \leq \boldsymbol{m} \leq 1$

b = strength of an alternate cause B of the effect(s), $0 \leq \boldsymbol{b} \leq 0.25$

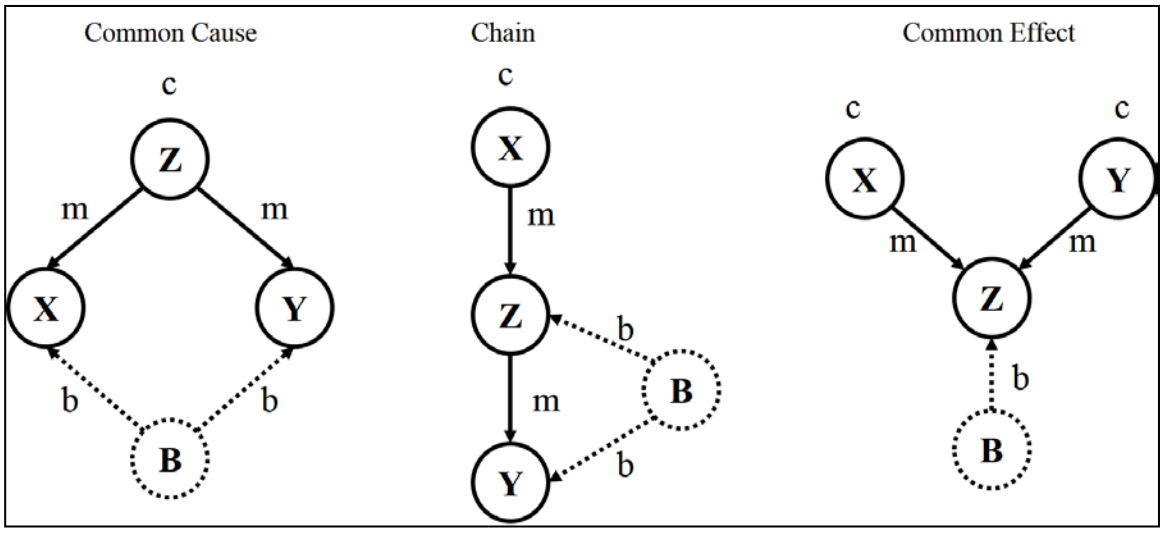


Fig. D1. Causal network structure defined by the normative causal graphical model (CGM). The arrows show the direction of causality. Dotted lines and nodes represent variables that were not part of the experimental setup taught to participants but are auxiliary variables assumed by the model to be part of the participants' mental construction of the causal system. (adapted from Rehder, 2014)

A value of 0.5 for ***c*** would imply a neutral view, that is, an equal probability of the presence or absence of the primary causes. A non-neutral view would be a biased view where reasoners may hold a bias towards the primary causes being present or absent (although there is no information in the experiments to suggest that any primary cause is more likely to be present or absent across trials). A neutral view is unbiased as it suggests that participants do not bring to the task any a priori assumptions about the presence of a primary cause, independent of the information provided in the experiments. A value of 1 for ***m*** would imply a deterministic causal relationship (i.e. the presence of a cause necessitates the presence of the effect), whereas any value less than one makes it a probabilistic

causal link. Note that alternate cause (***B***) is not part of the learned causal structure. It is an auxiliary variable, assumed to be constructed by the participants on the assumption that the link from the primary cause to the effects may not be deterministically necessary, that is, there may be alternate causes that influence the effects. Since this is not part of the actual structure taught to the participants, the parameter space in our implementation is restricted to smaller values (like the values tested by Rehder 2014).

We adapt this model for implementation within a Bayesian inference framework. A condensed version of the graphical model for the common cause CGM is shown in Figure D2. For the sake of exposition, a fully expanded version of the model is shown and described in Figure D3. The probabilities of y1 between two situations are translated into a relative probability using the softmax rule. For example, when comparing situations F and G, pFG denotes the relative probability of selecting F (e.g. 0.75 implies selecting F rather than G 75% of the time). This is calculated as

$$pFG = \frac{e^{\frac{logit(pF)}{\tau}}}{e^{\frac{logit(pF)}{\tau}} + e^{\frac{logit(pG)}{\tau}}} \quad (7)$$

and the same expression is used for the QP model. Finally, the actual choice score, i.e. the proportion of N questions of type FG where F was selected over G is modeled using a stochastic process Binomial (2N, pFG) / 2N.

In Figure D2, the probabilities of states A to H are represented as pV1 and pV2 (e.g. for the problem type AB, pV1 is the probability of state A and pV2 is the probability of state B), repeated over the t=1:5 different problem types (in this experiment, there are five different problem types, AB, BC, DE, FG and GH). pV represents the respective relative probabilities {pAB, pBC, pDE, pFG, pGH}. The detailed calculations for $pV1_t$ and $pV2_t$ are represented simply as $pV1_t$ = **CGM** (c_i, m_i, b_i, $V1_t$) and $pV2_t$ = **CGM** (c_i, m_i, b_i, $V2_t$). The following hierarchical prior construct follows, with the *'i'* sub-script denoting individual level parameters. The mean μ and standard deviation σ reflect the hyper parameters of the hierarchical distributions. Tau (τ) represents a temperature parameter for the softmax rule and is set to 1 in all the models considered. T(p,q) refers to truncation within the range [p,q]. The mean hyperparameter μ_m is drawn from the range [0.5,1] to reflect the fact that participants having being taught the causal relationship would likely infer a causal link greater than chance (0.5) levels. Figure D3 shows an expanded version of the model.

$\mu_m \sim$ Uniform (0.5,1)	$\mu_c \sim$ Uniform (0,1)	$\mu_b \sim$ Uniform (0,0.25)
$\sigma_m \sim$ Uniform (0.001,4)	$\sigma_c \sim$ Uniform (0.001,4)	$\sigma_b \sim$ Uniform (0.001,4)
$m_i \sim N(\mu_m, \sigma_m)_{T(0,1)}$	$c_i \sim N(\mu_c, \sigma_c)_{T(0,1)}$	$b_i \sim N(\mu_b, \sigma_b)_{T(0,0.25)}$

Conjunctive Model [CONJ]

The Conjunctive model is similar to the CGM but the probabilities of the target being present (i.e. y1) under each state (A to H) are calculated conjunctively rather than normatively. For instance, under situation F, the probability pF is calculated simply as p(cause does not exist and both effects exist) rather than the normative calculation p(cause does not exist and both effects exist) / (p(cause does not exist and one of the two effects exist) + p(cause does not exist and both effects exist)) for the CGM model shown in figure D3. Similar calculations are performed for all the states. This essentially implies that the states A-H are not treated as given (i.e. the probability of the given values of X and Z is not treated as 1), but the probability of y1 is evaluated in conjunction with the probability of the state of X and Z. See Rehder (2014) for further details of the CONJ model. The graphical Bayesian implementation (not shown) is similar to that for the CGM model (Figure 6), with the probabilities for the eight possible states (pV1 and pV2, where pV1, pV2 ϵ { pA, pB, pC, pD, pE, pF, pG, pH }) calculated using the CONJ rather than the CGM model.

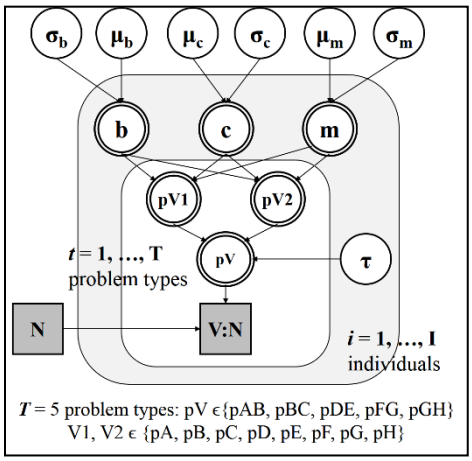


Fig. D2. Condensed graphical model for the normative causal graphical model [CGM]. The probabilities of states A to H are represented as pV1 and pV2, repeated over the T = 5 different problem types. pV represents the respective relative probabilities {pAB, pBC, pDE, pFG, pGH}. The calculations for $pV1_t$ and $pV2_t$ can be represented simply as $pV1_t$ = CGM (c_i, m_i, b_i, $V1_t$) and $pV2_t$ = CGM (c_i, m_i, b_i, $V2_t$), where CGM (c, m, b, V) is defined as per the detailed model. The notation V:N refers to the output of a binomial process with 2N trials, probability pV.

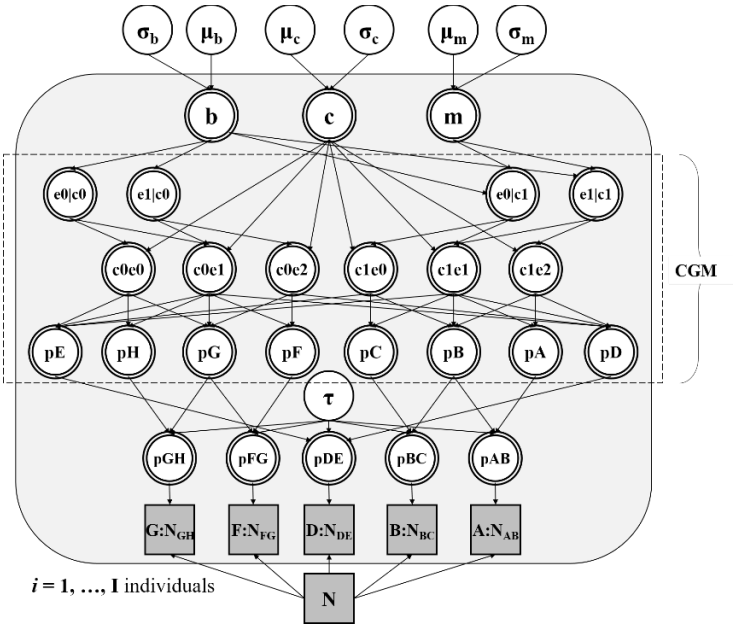


Fig. D3. Expanded graphical model for the normative causal graphical model [CGM]. Expanded graphical model for the CGM common cause network model (adapted from Rehder, 2014). In the figure, {e0, e1, e2} represent states with 0, 1 and 2 effects 'present', {c0, c1} represents states with the cause 'absent' and 'present' respectively. For example, (e1|c0) represents the independent probability of an effect being present conditional on the cause (Z) being absent, in this case being equal to ***b***. Similarly, c0e1 represents the joint probability of a state with the primary cause absent and one of the two effects present. This can be derived as p(c0)p(e1|c0)p(e0|c0), which is (1-c)b(1-b). These joint probabilities can then be used to calculate the probability of each of the states A to H. For instance, state F represents a state with the primary cause being absent z0, one effect being present x1 and the state of the second effect unknown. The probability of the second effect (Y) being present (denoted as pF) can be calculated as pF =

p(c0,e2) / (p(c0,e1) + p(c0,e2)) = b^2.(1-c) / ((1-b).b.(1-c) + b^2.(1-c)) = b. The notation nAB refers to the output of a binomial process with 2N trials, probability pAB.

Associative Model [ASSC]

The associative model (ASSC; Figure D4) posits that reasoners do not take into account the direction of causality, but that variables are associatively linked only by symmetric connections. Rehder (2014) models this as a Markov random field (an undirected graph with a Markov property, that describes the dependencies between variables), with the parameters a2 and a3 capturing the pairwise and three-way associative strengths between the variables. This model is common for all three network structures. The parameter a2 represents the strength of the pairwise association, that is, the strength of X-Z (or Y-Z) pair having the same state values (0 or 1). The parameter a3 represents the strength of the 3-way association, that is, the strength of X-Z-Y all having the same state value (0 or 1). The probability of any combination of values for X, Y and Z is obtained from this model as;

$$p(X_i, Y_j, Z_k) = Normalized\left(e^{-\left(f_2(X_i,Z_k) + f_2(Y_j,Z_k) + f_3(X_i,Z_k,Y_j)\right)}\right) \quad (8)$$

Here, f2(p,q) = a2 if p=q, 0 otherwise and f3(p,q,r) = a3 if p=q=r, 0 otherwise (see Rehder, 2014 for further details and properties of the model). The Bayesian graphical model (not shown) for implementation of the ASSC model is similar to the one shown for CGM (Figure D2), with $pV1_t$ = **ASSC** ($a2_i$, $a3_i$, $V1_t$) and $pV2_t$ = **ASSC** ($a2_i$, $a3_i$, $V2_t$). Hierarchical priors are set for the parameters a2 and a3 in the same way as the priors for the parameters of the CGM model, with the exception that a2 and a3 are constrained to the range [0, 3], since they reflect the relative strength and not probabilities (the range is similar to that explored in Rehder, 2014).

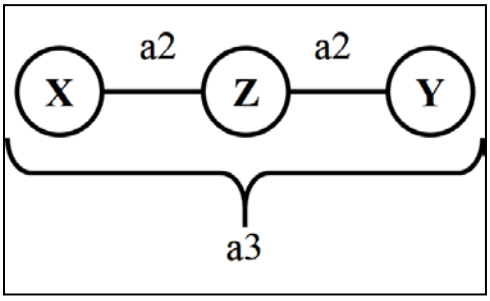


Fig. D4. Network structure defined by the Associative model (ASSC). The associative model does not assume any particular direction of causality (adapted from Rehder, 2014).

Specific Shared Disabler Model [DISAB]

Figure D5 shows the specific shared disabler model adapted from Rehder (2014). The key difference between the normative CGM and the disabler (DISAB) model is that the latter introduces an additional auxiliary variable, which is a correlated external influence on both X and Z. This disabling mechanism adds a link between the effects, allowing inferences to be influenced by otherwise normatively independent variables (e.g. between Y and X in the case of the common cause structure) without a violation of the Causal Markov condition (see Rehder (2014) for a detailed discussion and elaboration of the model).

The parameters c, m, and b have the same interpretation as the CGM and CONJ models. Two additional free parameters are *d*, the prior probability of the shared disabler W (range [0, 1]) and *dm*, the power of the causal link from the shared disabler W to the effects (range [0, 1]). Note that similar to *B*, the shared disabler *W* is not part of

the taught causal structure, but an auxiliary variable assumed to be constructed by the participants on the assumption that there may exist an interactive causal disabling influence that when present, probabilistically disables the primary causal mechanisms (e.g. from Z to X and Y in the common cause structure). The Bayesian graphical model (not shown) for implementation of the DISAB model is similar to the one shown for CGM (Figure D2), with $pV1_t$ = **DISAB** (m_i, c_i, b_i, d_i, dm_i, $V1_t$) and $pV2_t$ = **DISAB** (m_i, c_i, b_i, d_i, dm_i, $V2_t$).

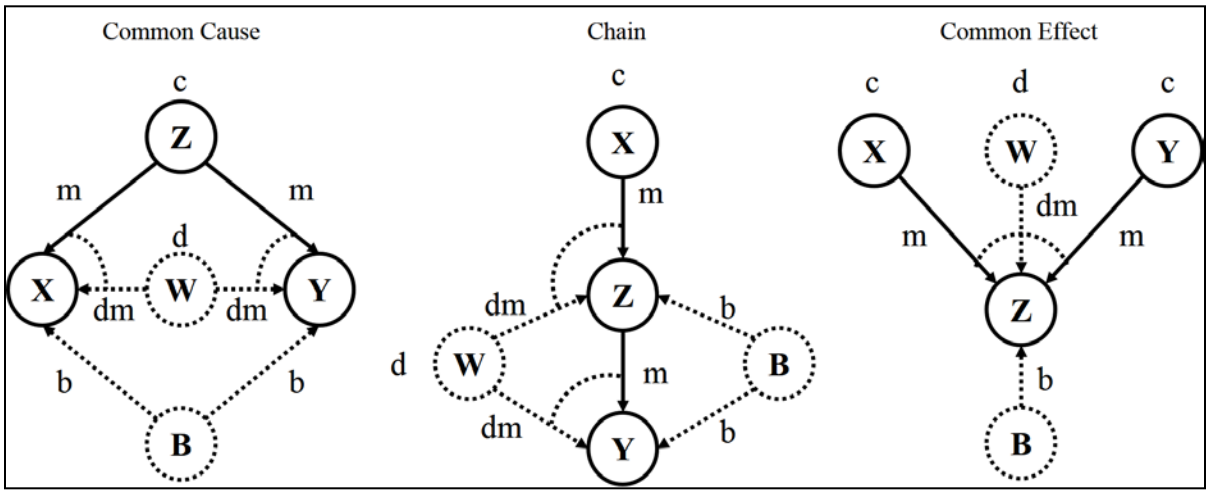


Fig. D5. Causal network structure defined by the specific shared disabler model (DISAB). The arrows show the direction of causality. Dotted lines and nodes represent variables that were not part of the experimental setup taught to participants but are auxiliary variables assumed by the model to be part of the participants' mental construction of the causal system (adapted from Rehder, 2014)

E. Model metrics (correlation, bias, RMSE) by causal vs non-causal reasoners

Common Cause	Causal Reasoners			Associative Reasoners		
	Correlation	Bias	RMSE	Correlation	Bias	RMSE
CGM	0.75	-0.020	0.122	0.30	-0.265	0.338
CONJ	0.32	-0.062	0.251	0.36	-0.248	0.370
ASS	0.78	0.028	0.121	0.36	-0.133	0.235
DISAB	0.79	0.012	0.113	0.53	-0.156	0.238
WTM	0.75	0.024	0.124	0.54	-0.153	0.225
QPN	**0.91**	**-0.003**	**0.08**	0.88	-0.07	0.11
QPC	**0.91**	-0.007	**0.08**	**0.93**	**-0.04**	**0.08**
Chain	Causal Reasoners			Associative Reasoners		
	Correlation	Bias	RMSE	Correlation	Bias	RMSE
CGM	0.73	-0.046	0.148	0.29	-0.271	0.342
CONJ	0.42	**-0.001**	0.218	0.12	-0.162	0.314
ASS	0.75	0.008	0.136	0.45	-0.133	0.222
DISAB	0.63	0.035	0.167	0.60	-0.073	0.161
WTM	0.80	0.021	0.123	0.45	-0.150	0.231
QPN	0.86	-0.02	0.10	0.78	-0.09	0.15
QPC	**0.91**	-0.02	**0.09**	**0.91**	**-0.04**	**0.09**
Common Effect	Causal Reasoners			Associative Reasoners		
	Correlation	Bias	RMSE	Correlation	Bias	RMSE
CGM	0.43	0.019	0.214	0.15	-0.258	0.334
CONJ	0.41	0.027	0.245	0.39	-0.108	0.269
ASS	0.48	0.116	0.248	0.43	-0.098	0.218
DISAB	0.53	0.044	0.207	0.25	-0.250	0.325
WTM	0.73	0.062	0.187	0.61	-0.132	0.217
QPN	**0.88**	0.03	**0.11**	**0.93**	-0.07	0.11
QPC	**0.88**	**0.02**	**0.11**	**0.93**	**-0.05**	**0.10**

F. Model fit by heuristic models (CGM, CONJ, ASSC, DISAB) by causal vs non-causal reasoners

One of the purposes of this work is to understand in more detail the difference between causal and associative reasoners Rehder (2014) postulated. To this end, we undertook a comparison of aggregate (across participants) means of the posterior predictive choice responses predicted by the four different heuristic strategies, separately for causal and non-causal reasoners (as identified in Rehder, 2014), for each of the five problem types (AB, BC, DE, FG and GH). These are shown in figures F1, F2, F3 for the three types of network structures.

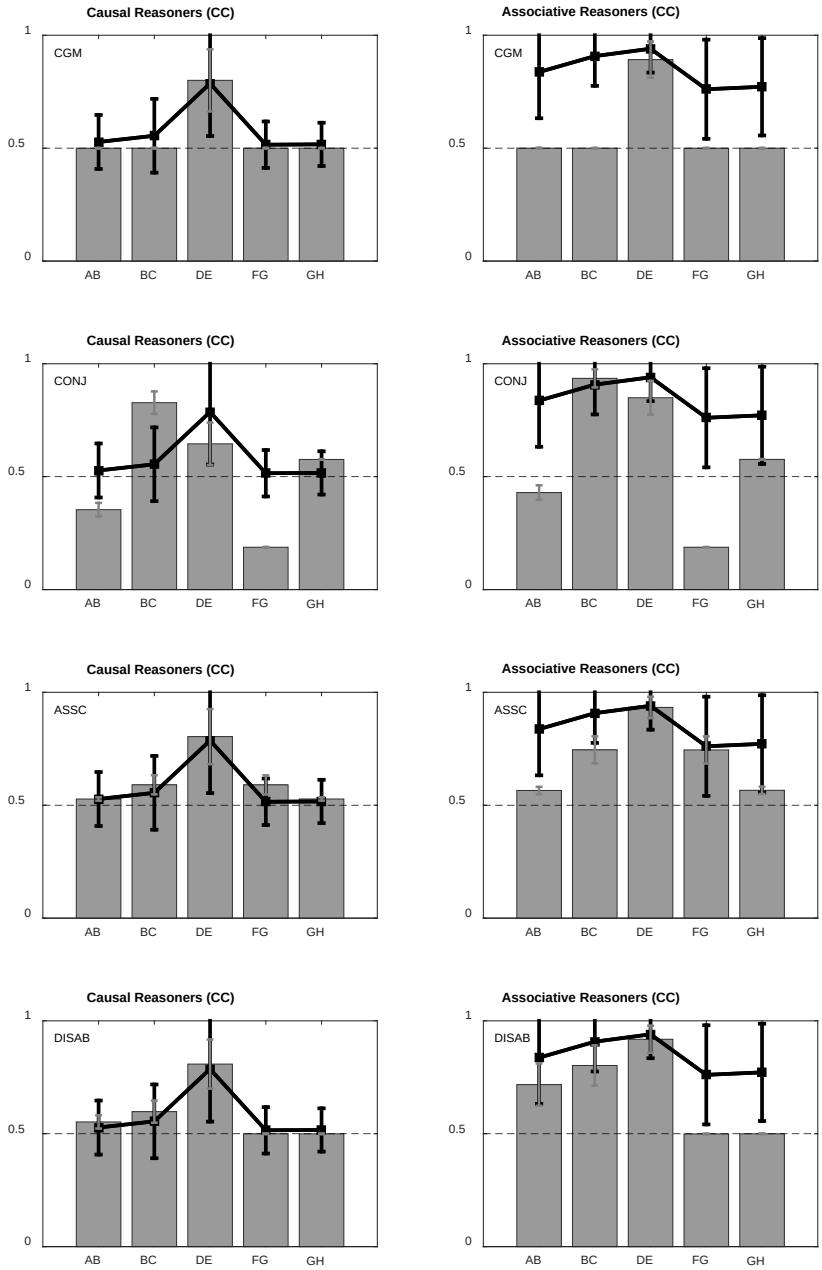


Fig. F1: Mean posterior predictive vs actuals by heuristic models (COMMON CAUSE)

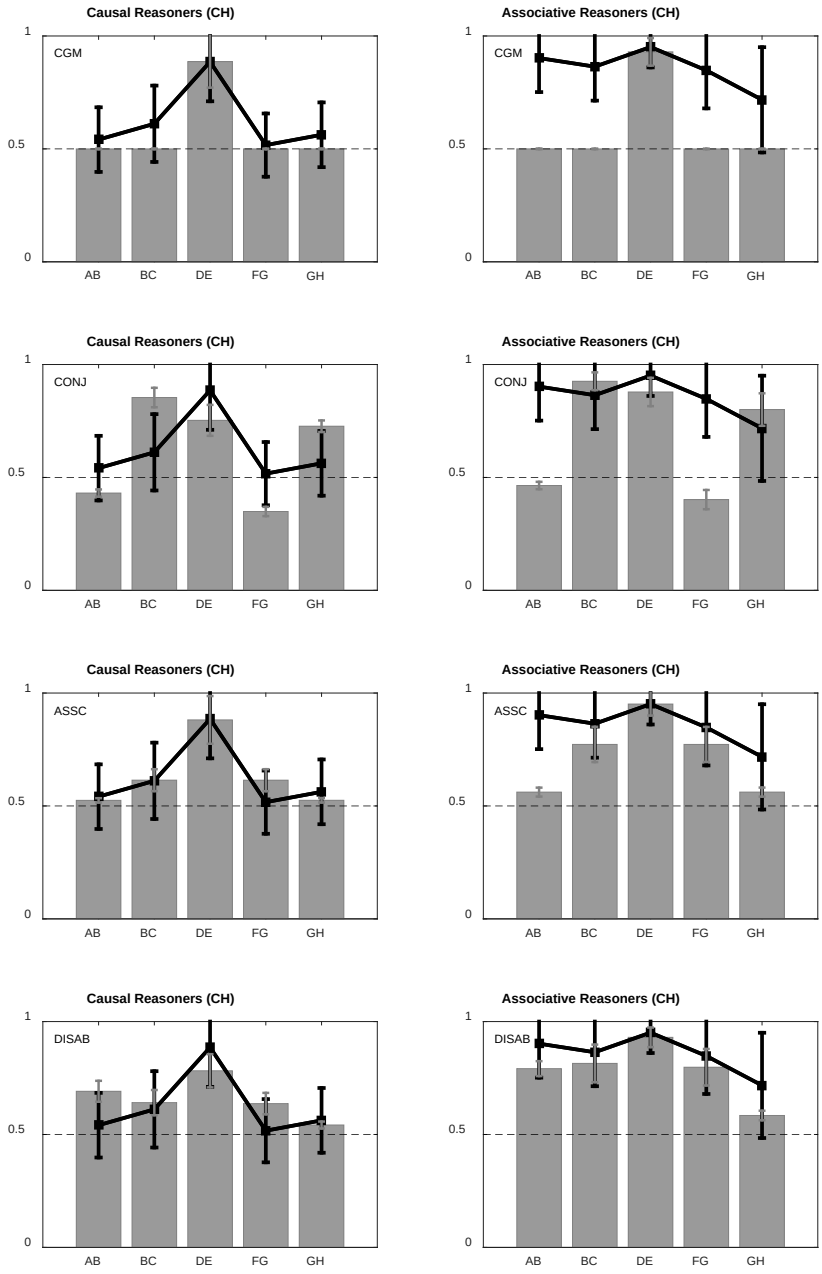


Fig. F2: Mean posterior predictive vs actuals by heuristic models (CHAIN)

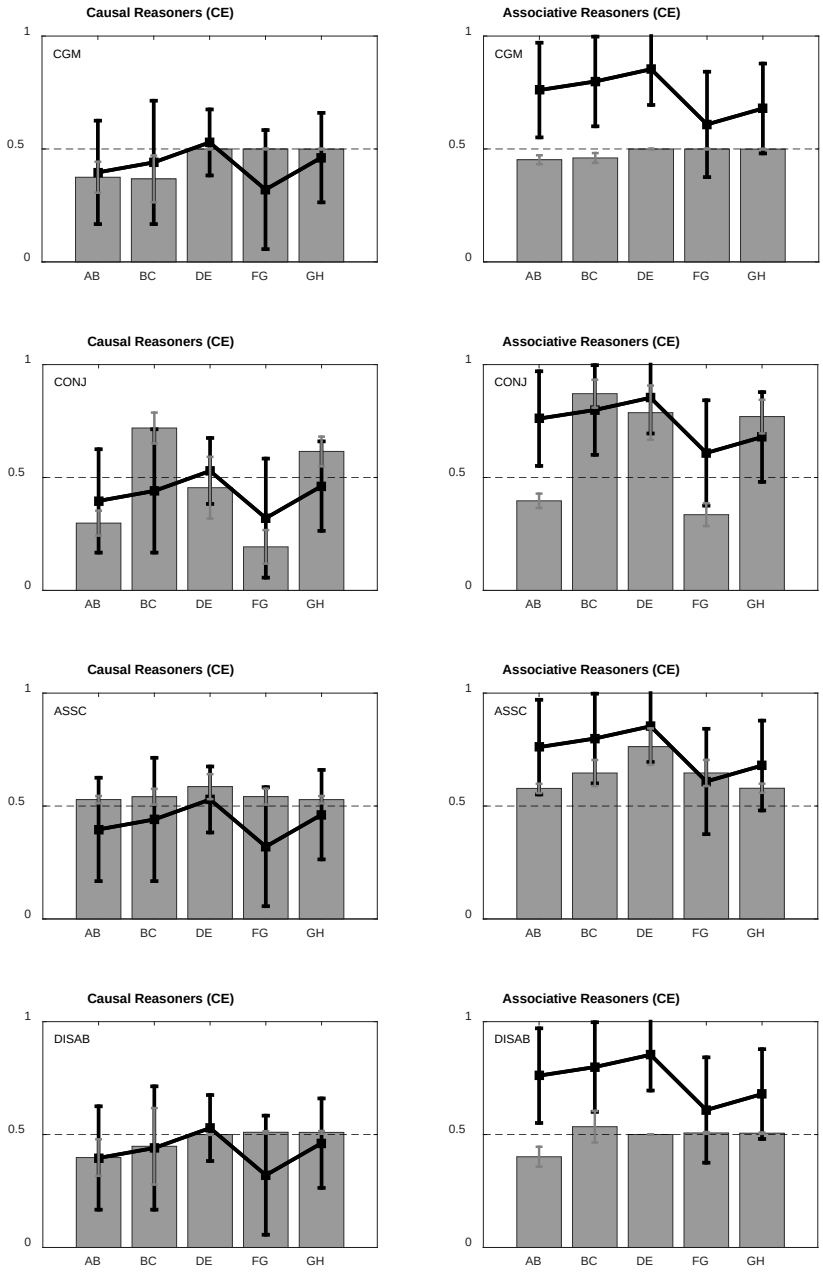


Fig. F3: Mean posterior predictive vs actuals by heuristic models (COMMON EFFECT)

G. Inferred strategy weights for the WTM (dataset 1; comparative judgments)

Figure G1 shows the mean inferred weights for the WTM model for each individual participant and Table G2 summarizes these values across participants. Associative reasoners show higher weights for the ASSC model than causal reasoners, causal reasoners show higher weights for the CGM model, as expected, and the CONJ model is shown to have a slightly higher weight in the CE structure compared to the value in Rehder (2014). However, the mean weights inferred using the hierarchical Bayesian approach at an individual level differ from the aggregate weights reported in Rehder (2014), which were concentrated towards the CGM and ASSC models for the causal and associative reasoners respectively. The inferred weights for most participants show a reasonable combination of all four strategies, rather than a significant preference for a single strategy for each participant, especially for causal reasoners. This seems to be a result of two aspects. First, these strategies often make overlapping predictions and can account for normative behavior in a similar manner. Second, when modeling individual differences, both the parameters of the individual strategies and the weights of these strategies can vary at an individual level, providing a large number of degrees of freedom to the model to account for behavior. For example, the ASSC model shows a higher than expected weight (approximately 0.24) for causal reasoners in the CC and CH networks. However, when applied individually, the mean inferred *a3* parameter for causal reasoners is much lower than that for associative reasoners (see the third row in figure 13). Hence, the weights of the models are not directly comparable, since their influence also depends on the underlying parameter values.

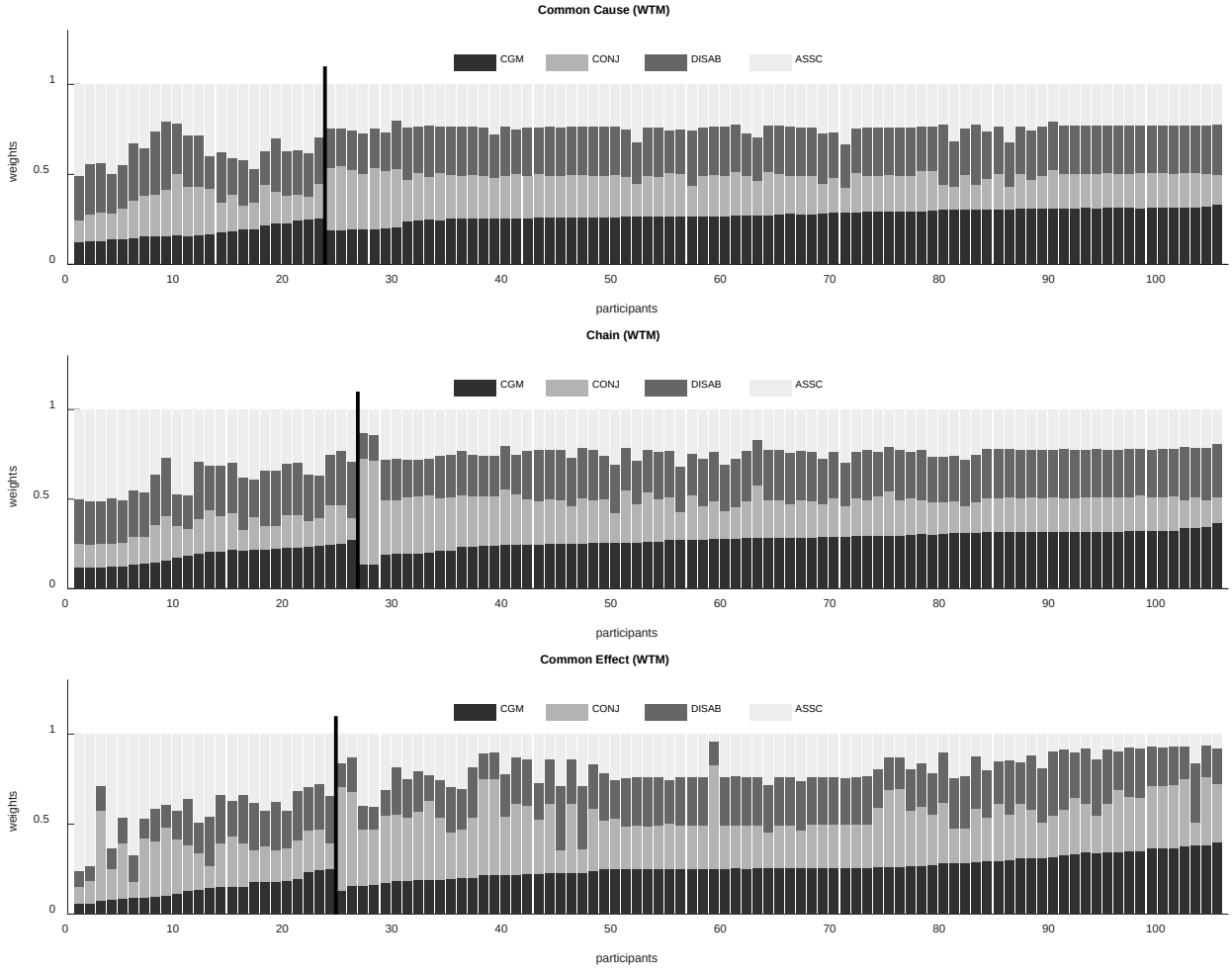


Figure G1: The mean posterior weights of the 4 strategies inferred as per the WTM model for each individual participant. The participants to the left of the black line in each plot are the non-causal reasoners and to the right are the causal reasoners (as per Rehder, 2014 classification).

Table G2: Summary analysis (mean and standard deviation) of the inferred weight parameters in the WTM model. Numbers in bold highlight large differences between causal and associative reasoners.

WTM model		Common Cause		Chain		Common Effect	
		mean	std	mean	std	mean	std
w_{CGM}	Overall	0.26	0.05	0.26	0.06	0.24	0.08
	Causal	**0.28**	0.03	**0.28**	0.04	**0.26**	0.06
	Associative	**0.18**	0.04	**0.19**	0.05	**0.14**	0.06
w_{CONJ}	Overall	0.21	0.05	0.21	0.07	0.28	0.10
	Causal	0.22	0.05	0.23	0.07	**0.30**	0.09
	Associative	0.19	0.06	0.17	0.04	**0.23**	0.09
w_{ASSC}	Overall	0.27	0.07	0.27	0.08	0.24	0.14
	Causal	**0.24**	0.02	**0.24**	0.03	**0.19**	0.08
	Associative	**0.37**	0.08	**0.38**	0.09	**0.44**	0.14
w_{DISAB}	Overall	0.26	0.03	0.26	0.03	0.23	0.06
	Causal	0.26	0.02	0.26	0.03	0.25	0.05
	Associative	0.26	0.05	0.27	0.04	0.19	0.07

H. Cluster level rotation parameters (dataset 1; comparative judgments)

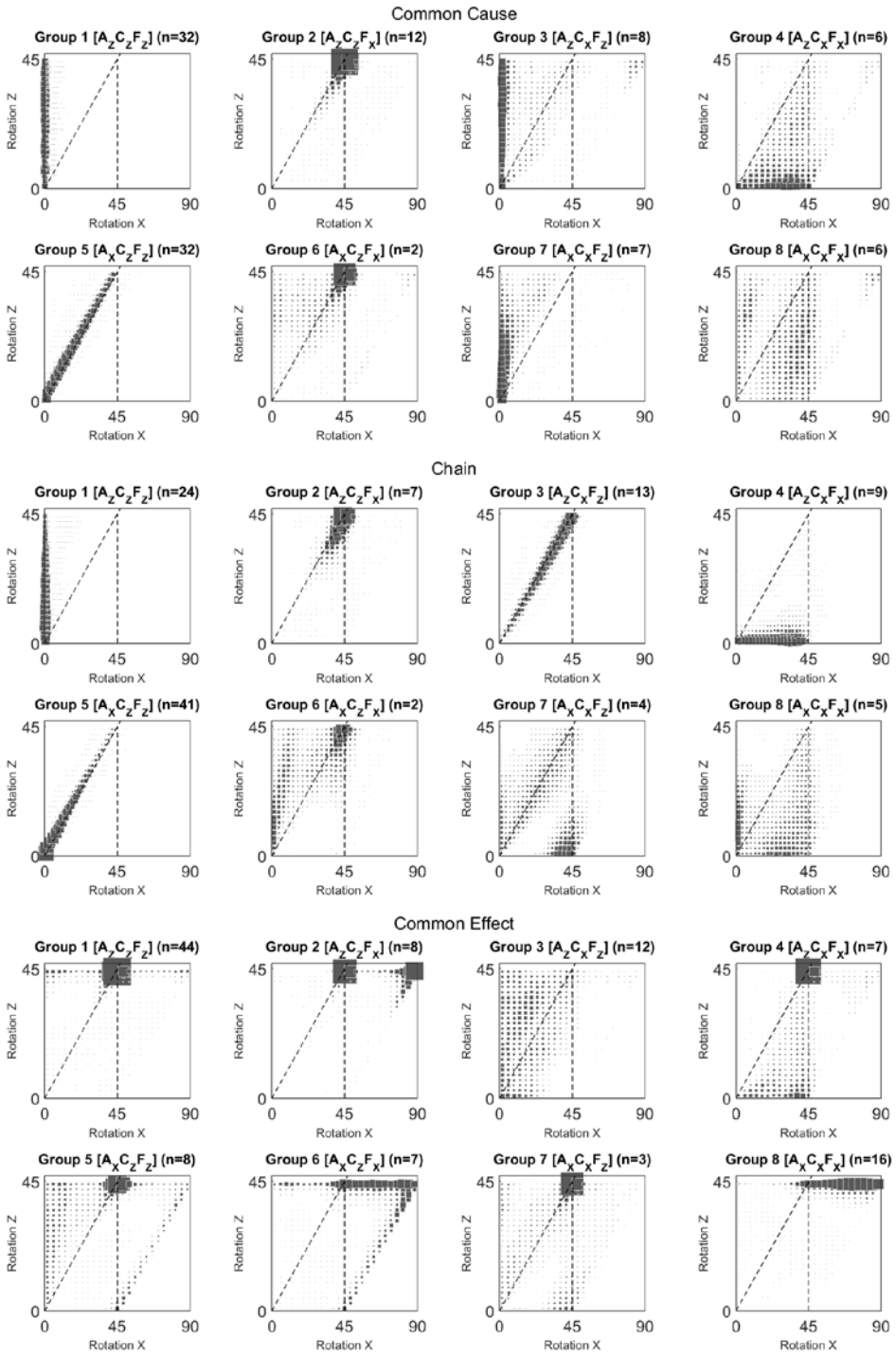


Fig. H1. Joint posterior probability density of the rotation parameters of the QPC model. The parameters are continuous but have been binned for this plot. The size of squares shows the value of the joint posterior probability density of the parameters within each cluster. Each cluster also shows the preferred projection orders for that cluster, and the number of participants for which this was the model cluster.

I. RMSE between each cluster and normative behavior

To provide an overview of normative behavior, we computed root mean square error (RMSE) between mean proportions for each comparison (five comparisons per network) and normative point values (based on CGM predictions, as in Rehder, 2014). In Table I1, we show results for 4/5 comparisons (we excluded comparison GH for brevity and because GH results are less diagnostic concerning individual differences); the overall RMSE for each cluster is computed across all 5 comparative judgments including GH.

Overall, for both CC and CH networks, participants who can focus on the immediate variable (normative) seem also able to incorporate other aspects of normative behavior, notably independence (CC) and compression (CH). Focus on the immediate variable and assumptions regarding symmetry or compression are interesting since they could influence behavior at the level of overall assumptions about the problem (e.g., regarding symmetry, a participant may think "I have been told that the variables are independent and conditionals should reflect this"). However, in the case of the DE comparisons, normative behavior is a matter of finer tuning of rotation parameters (i.e., assumptions regarding the relatedness of the variables), to achieve the normative effect, and many participants are unable to do this. The picture for CE is somewhat different, with normative behavior apparently easier to achieve in DE comparisons, but evidenced lack of sensitivity in relation to discounting effects.

Table I1: RMSE calculations based on the cluster mean proportions and normative point values based on CGM. High RMSE indicates high deviation from normative behavior. Values close to zero indicate normative-like behavior[17]

			RMSE vs normative behavior			
Cluster	Conflict	Symmetry	**Conflict (BC, FG)**	**SYM (AB)**	**DE**	**overall**
1. A_{Z1}-C_{Z1}-F_{Z0}	Immediate	No	0.09	0.05	0.28	0.15
2. A_{Z1}-C_{Z1}-F_{X1}	Present	No	0.08	0.16	0.43	0.21
3. A_{Z1}-C_{X0}-F_{Z0}	Absent	No	0.37	0.14	0.27	0.28
4. A_{Z1}-C_{X0}-F_{X1}	Distant	No	0.39	0.48	0.25	0.38
5. A_{X1}-C_{Z1}-F_{Z0}	Immediate	Yes	0.13	0.14	0.05	0.12
6. A_{X1}-C_{Z1}-F_{X1}	Present	Yes	0.40	0.36	0.20	0.35
7. A_{X1}-C_{X0}-F_{Z0}	Absent	Yes	0.38	0.47	0.09	0.35
8. A_{X1}-C_{X0}-F_{X1}	Distant	Yes	0.39	0.25	0.40	0.35

			RMSE vs normative behavior			
Cluster	Conflict	Compression	**Conflict (BC, FG)**	**CMP (AB)**	**DE**	**overall**
1. A_{Z1}-C_{Z1}-F_{Z0}	Immediate	No	0.09	0.10	0.20	0.12
2. A_{Z1}-C_{Z1}-F_{X1}	Present	No	0.11	0.05	0.35	0.18
3. A_{Z1}-C_{X0}-F_{Z0}	Absent	No	0.29	0.15	0.19	0.22
4. A_{Z1}-C_{X0}-F_{X1}	Distant	No	0.41	0.50	0.13	0.39
5. A_{X1}-C_{Z1}-F_{Z0}	Immediate	Yes	0.15	0.22	0.05	0.15
6. A_{X1}-C_{Z1}-F_{X1}	Present	Yes	0.34	0.44	0.06	0.34
7. A_{X1}-C_{X0}-F_{Z0}	Absent	Yes	0.43	0.31	0.32	0.39
8. A_{X1}-C_{X0}-F_{X1}	Distant	Yes	0.47	0.29	0.05	0.35

			RMSE vs normative behavior			
Cluster	Conflict	Discounting	**Conflict (FG)**	**DISC (AB, BC)**	**DE**	**overall**
1. A_{Z1}-C_{Z1}-F_{Z0}	Immediate	Ignore alternate cause	0.12	0.45	0.14	0.30
2. A_{Z1}-C_{Z1}-F_{X1}	Distant	Ignore alternate cause (monotonicity violated)	0.49	0.49	0.15	0.40
3. A_{Z1}-C_{X0}-F_{Z0}	Immediate	Anti-discounting (monotonicity violated)	0.23	0.76	0.40	0.54
4. A_{Z1}-C_{X0}-F_{X1}	Distant	Anti-discounting	0.37	0.84	0.43	0.61
5. A_{X1}-C_{Z1}-F_{Z0}	Immediate	Anti-discounting (weak positive product synergy)	0.27	0.64	0.28	0.45
6. A_{X1}-C_{Z1}-F_{X1}	Distant	Discounting (strong negative product synergy)	0.39	0.35	0.05	0.31
7. A_{X1}-C_{X0}-F_{Z0}	Immediate	Anti-explaining away (weak positive product synergy)	0.32	0.76	0.16	0.51
8. A_{X1}-C_{X0}-F_{X1}	Distant	Explaining away (strong negative product synergy)	0.48	0.23	0.14	0.29

[17] Note, even when behavior indicates focus on the immediate variable, which is normative, there may be a non-zero RSME value partly because of the probabilistic way in which participants were assigned to clusters, and because of deviation of participant behavior from exact point values from the CGM, even if the participant's behavior is qualitatively normative.

J. Plausibility Check of absolute probabilities for comparative judgments

Given a certain state of the variable X, higher absolute probability judgments when Z is present (z1) rather than absent (z0) for the common cause and chain networks are cognitively more plausible. Figure J1 plots corresponding conditional QPC predicted probabilities (the diagonal indicates equality of the conditionals). The first two panels in Figure J1 show that, as expected, the intermediate absolute probability judgments are generally higher (to the right of the diagonal) when Z is present compared to when they are absent, for specific values of X. The third panel concerns the common effect structure. When the value of X is known (either 0 or 1), the mean absolute predicted judgments tend to be similar (lie mostly along the diagonal) for most of the participants, suggesting that X (the value that is the same in the 2 situations compared) rather than Z (which is different between the two situations compared) is a key factor for judgments, which is highly plausible given that in the common effect structure the alternate cause (X) may play a strong role, especially when its value is known. When the value of the alternate cause (X) is unknown (data pattern vertical to the diagonal), the judgments for most the participants seems to be higher when Z is present, but there is also a reasonable share of participants where the opposite is true. Overall, the intermediate absolute probability judgments show a high level of cognitive plausibility.

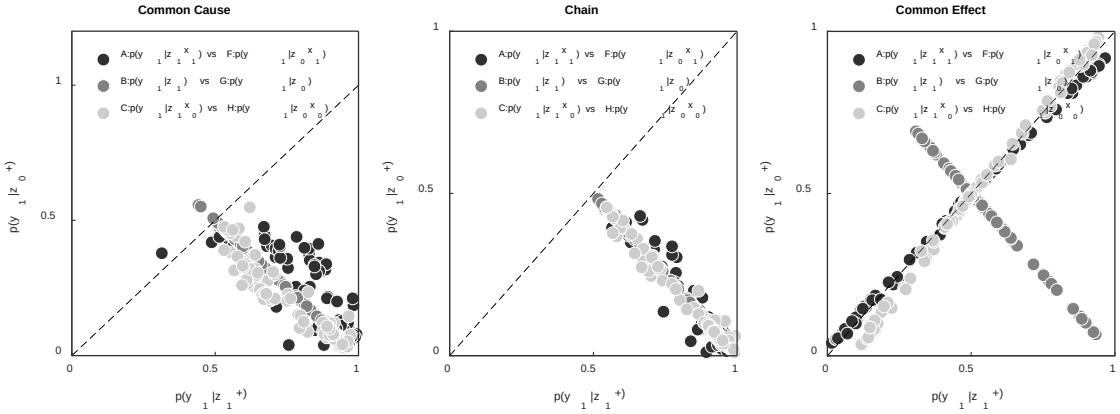


Fig. J1. Plausibility of intermediate absolute judgments (dataset 1): Comparison of intermediate absolute probability judgments generated by the clustered QPC model. Each dot represents the mean intermediate probability judgments for an individual. The x-axis plots the values when Z is present and the y-axis when Z is absent. The comparisons are made for comparable values of X (present, absent, unknown). Values to the right of the diagonal show judgments that are higher when Z is present compared to when Z is absent.

K. Calculation of Bayes Factors using the product-space method

To calculate Bayes factors between the QP and WTM models, we use the product space method (Lodeyckx et al, 2011). The schematic implementation is provided below:

Model Index
M_i ~ Bernoulli ($Model_{prior}$)
$Model_i = M_i + 1$
Model Likelihood
$Data_i$ ~ Binomial($Probability_i[Model_i]$)
$Probability_i[1] = Probability_i[WTM]$
$Probability_i[2] = Probability_i[QP]$
Model 1: WTM
$Probability_i[WTM] = WTM(parameters_W[Model_i])$
$parameters_W$ [1] ~ $parameters_W$.Prior
$parameters_W$ [2] ~ $parameters_W$.PseudoPrior
Model 2: QP
$Probability_i[QP] = QP(parameters_Q[Model_i])$
$parameters_Q$ [1] ~ $parameters_Q$.PseudoPrior
$parameters_Q$ [2] ~ $parameters_Q$.Prior

Here, $parameters_W$ and $parameters_Q$ refer to the vector of all parameters of the WTM and QP models respectively. $WTM(parameters_W[Model_i])$ represents the full WTM model, and generates a probability for each choice based on parameters $parameters_W[Model_i]$. The full QP model is represented as $QP(parameters_Q[Model_i])$, and generates a probability for each choice based on parameters $parameters_Q[Model_i]$. For each individual ***i***, each iteration of the MCMC sampling selects either $Model_i = 1$ (WTM) or $Model_i = 2$ (QP). If the WTM model is selected, the parameters $_W[1]$ reflect a draw from the actual priors / hyper-priors. On the other hand, since the QP model is disconnected from the data, the $parameters_Q$ [1] reflect a draw from a pseudoprior. The opposite happens if the QP model is selected. For all the hyper-parameters of the WTM and QP models, the pseudo priors are normal distributions, and the parameters of these normal distributions are estimated by first running each of the two models in separate runs (as recommended in Lodeyckx et al, 2011) and using the summary statistics of the resulting MCMC samples, to construct the normal pseudopriors. We highlight that suitable pseudopriors are important to ensure good mixing and convergence of the MCMC sampling, but do not affect the resulting Bayes factor calculation. Figure 1 shows an example of the distribution of MCMC samples for a single parameter of the WTM model when run separately, and the normal approximation of this using summary statistics (in line with recommendations by Lodeyckx et al (2011), that pseudopriors be chosen from a known family of probability distributions) which is then used as the pseudoprior. Finally, to calculate the Bayes factors, we use 5 different priors $\mathbf{Model_{prior}}$ ranging from 0.01 to 0.99 in favor of each model. For each individual, we select the Bayes factor to report based on the prior that results in the maximum model switches and best mix of model activation (see section 4.3 in Lodeyckx et al, 2011).

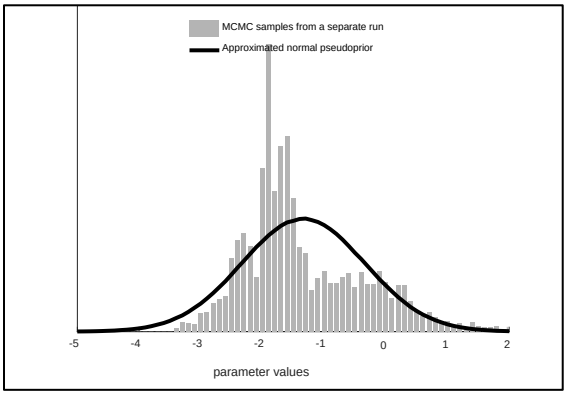


Figure K1: Example of pseudoprior calculation of a single parameter of the WTM model based on MCMC samples obtained from a separate run of the model.

L. Cluster level rotation parameters (dataset 2; absolute judgments)

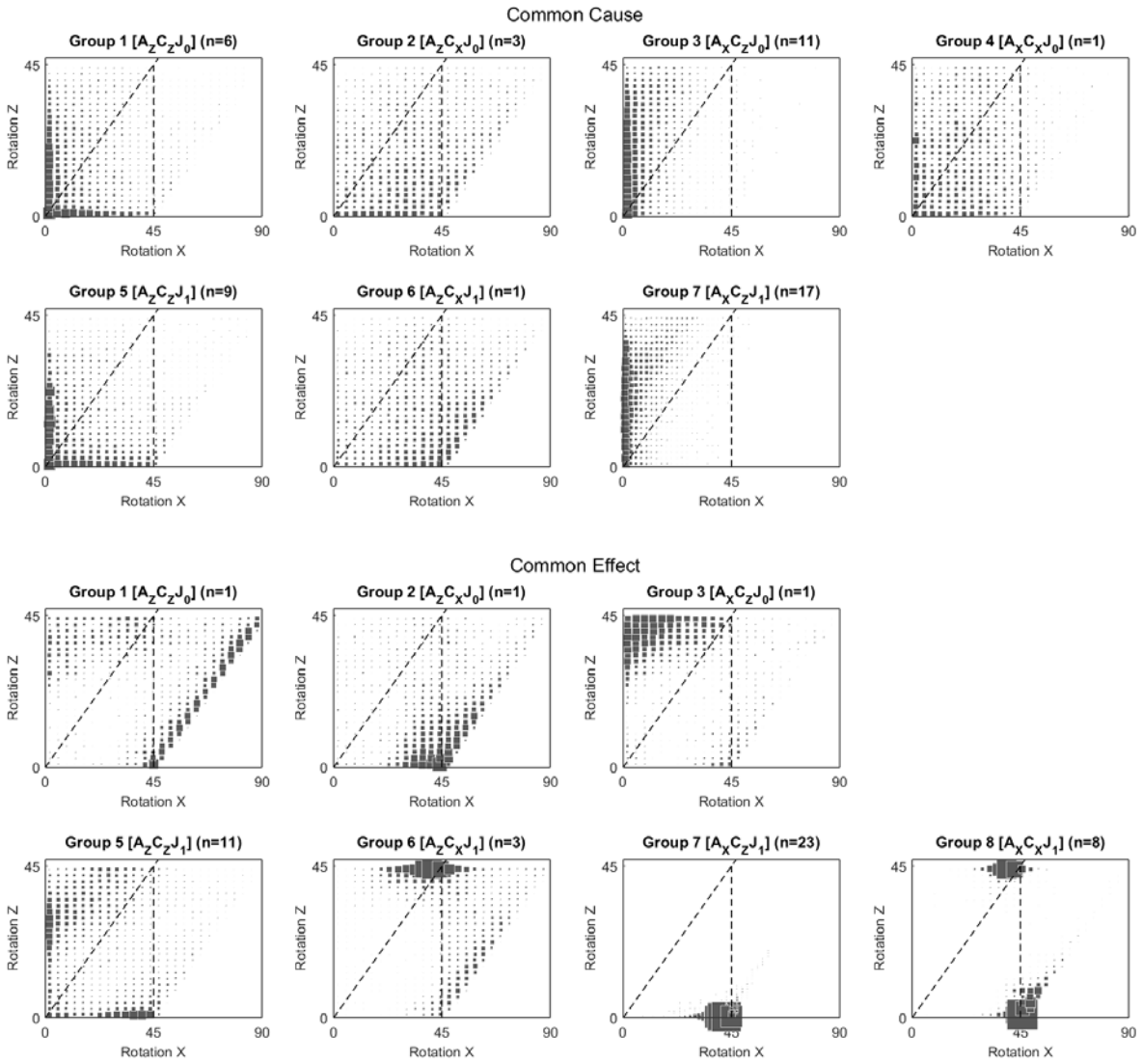


Fig. L1. Joint posterior probability density of the rotation parameters of the QPC model. The parameters are continuous but have been binned for this plot. The size of squares shows the value of the joint posterior probability density of the parameters within each cluster. Each cluster also shows the preferred projection orders for that cluster, and the number of participants for which this was the model cluster.

M. Parameter and data recovery for the QPC model

We use the range of parameters inferred from human data to simulate new data using the QPC model, then conduct a parameter and data recovery exercise using the QPC model. Figure M1 shows the mean recovered rotation parameters and mean data (probabilities in the range 0 to 1) analogous to the comparative judgment task. Both parameters and data show reliable and robust recovery, with strong correlations (0.93-0.94 for the parameters and 0.96 for the data). This confirms the identifiability and reliability of the QPC model.

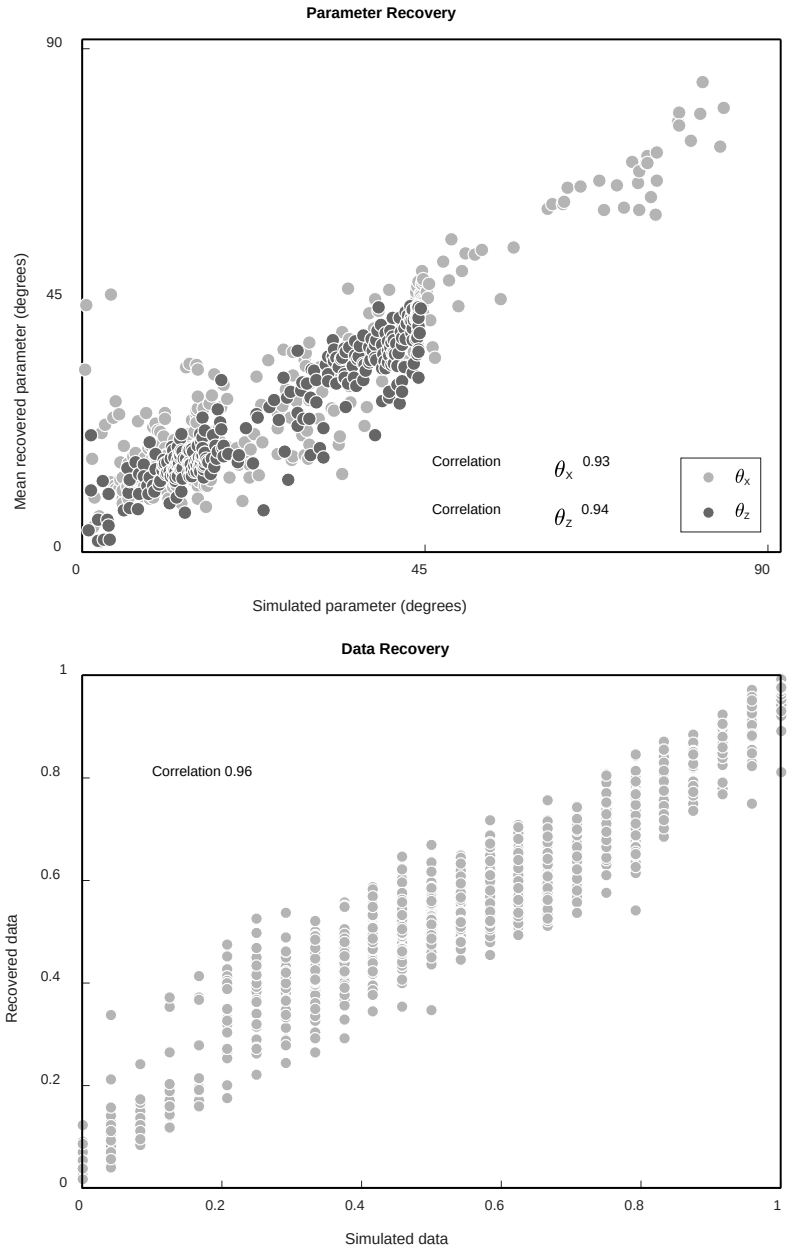


Fig. M1. Recovery of simulated rotation parameters and generated data by the QPC model

N. Using the QPC model to recover data generated by the WTM model

We use the WTM model to simulate data analogous to the comparative judgment task, and attempt recovery of this data using the QPC model. Table N1 shows how well the posterior predictive data generated by the QPC model data correlates with the input data to the model. The columns show the three sources of input data – based on the simulated WTM and QPC models as well as the actual human data. The correlation is robust and reliable for both QPC and human data, but is much weaker for the data simulated by the WTM model, showing that there is a significant combination of data points produced by the WTM model that cannot be effectively captured by the QPC model. This is especially true for comparisons AB and GH. This shows that the improved fit is not simply a matter of the QPC model having significantly greater flexibility.

Table N1: Correlation between posterior predictive of the QPC model and the input data for each comparison situation (AB, BC, DE, FG, GH). The columns show the three sources of input data – based on data simulated by the WTM model, by the QPC model, and the actual human data.

Comparison	WTM simulated	QPC simulated	Human data
AB	**0.46**	0.92	0.90
BC	0.73	0.98	0.93
DE	0.92	0.95	0.92
FG	0.81	0.96	0.93
GH	**0.49**	0.90	0.90
All	**0.81**	**0.96**	**0.93**