Causal Information Seeking

Brian N. Yin (bny212@nyu.edu)
Department of Psychology, New York University

Benjamin R. Scoblionko (bscoblionko@gmail.com)

Bob Rehder (br35@nyu.edu)
Department of Psychology, New York University

Abstract

How do people's causal knowledge influence how they seek information? The current work tasks participants with choosing to observe disease symptoms in a setting where they know a disease's etiology and related symptoms. We use causal graphical models (CGMs) to formalize their causal knowledge of the disease, and find that people tend to use their expected information gain, computed over their CGM-generated probability beliefs, to search for information in causal settings.

Keywords: Bayesian decision theory; information search; causal reasoning; causal graphical models

Introduction

How does a person's causal knowledge influence how they seek information? A doctor chooses a test for a patient presenting with a cough and chest pains, a cognitive psychologist designs an experiment to probe a popular theory, a child suffering from afternoon tummyaches asks the lunch lady what's in the chocolate cake. In both specialized and everyday contexts, we reason forward from causes to effects (tree nuts cause allergic reactions) and backward from effects to causes (from a patient's symptoms to a doctor suspects bronchitis from their patient's symptoms), and knowledge of what causes what influences how we search for explanations.

Information search has been studied extensively across cognitive psychology and underlies the research that's nominally about how people test hypotheses to discover general rules and regularities (Klayman & Ha, 1987; Markant & Gureckis, 2014; Oaksford & Chater, 1994; Wason, 1968) and about how they seek evidence to infer the categories of given exemplars and the underlying causes of particular events (Baron, Beattie, & Hershey, 1988; Eddy, 1982; Nelson, McKenzie, Cottrell, & Sejnowski, 2010; Rehder & Hoffman, 2005). However, whereas the research to date has generally involved participants reasoning with statistical information relating evidence and hypotheses (typically, the conditional probability of evidence given the presence of each hypothesis), reasoning in real-world scenarios often uses a reasoner's causal knowledge about how evidence relates to hypotheses.

A large body of empirical work in the last two decades, inspired by formalisms from artificial intelligence (Glymour, 1998; Pearl, 2000), has shown that people employ their causal knowledge during reasoning (Holyoak, Lee, & Lu, 2010; Kemp, Shafto, & Tenenbaum, 2012; Rehder & Burnett, 2005), learning (Cheng, 1997; Griffiths & Tenenbaum, 2005), decision making (Hagmayer & Sloman, 2009), and classification tasks (Rehder, 2003; Rehder & Kim, 2010), and that their behavior on these tasks can be explained by *causal graphical models* (CGMs) of their knowledge.

The present work will investigate exactly how people's causal knowledge influences how they choose what information to acquire in the domain of medical diagnosis, using CGMs to model their causal knowledge as a structure over which they search for information. We gave participants network diagrams that displayed a disease's etiology and related symptoms, and analyzed how participants' conditional probability judgements (e.g., their belief in the chance the disease gave rise to one of its characteristic symptoms) predicted their choices in an information-seeking task (their choice of what observation to make to best diagnose a hypothetical patient).

The Value of Information

Following Nelson et al. (2010), we quantify the value of an observation as its *expected utility*—the usefulness of the possible outcomes of the observation, each weighted by its probability and aggregated into a single value representing the observation's expected usefulness. We can express the expected utility $U(A)$ of observing an alternative as

$$U(A) := \sum_{a \in A} P(a)u(a) \tag{1}$$

where A is the random variable we're observing, $P(a)$ represents the probability of a specific value a of an observation, and $u(a)$ represents the utility of observing A to be a—the value of an observation coming out a specific way. We will test four utility functions, each corresponding to a different theory of how people value information.

Information gain (IG) (Lindley, 1956) treats the value of an observation as the degree to which it reduces uncertainty (reduces entropy of the hypothesis space):

$$u_{IG}(a) := \left[-\sum_{d \in D} P(d|a) \log P(d|a) \right] - \left[-\sum_{d \in D} P(d) \log P(d) \right] \tag{2}$$

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

D is a random variable[1] who's value we're interested in inferring from observing A, the first parenthetical represents our uncertainty about D's value after observing $A = a$, and the second parenthetical represents our uncertainty about D's value before observing A.

Probability gain (PG) (Baron & Hershey, 1988) treats the value of an observation as the increase in our probability of guessing D's value, given that we're guessing greedily:

$$u_{PG}(a) := \max_{d \in D}[P(d|a)] - \max_{d \in D}[P(d)] \tag{3}$$

where the first max expression is the chance we infer D's value correctly after observing $A = a$ and the second max expression is the chance we infer correctly before observing A.

Impact (Imp) (Klayman & Ha, 1987) assigns the value of an observation to be the summed absolute changes from prior to posterior beliefs over D's value:

$$u_{\text{Imp}}(a) := \sum_{d \in D} |P(d|a) - P(d)|. \tag{4}$$

Certainty gain (CG) (Baron & Hershey, 1988) assigns the value of an observation to be the number of hypotheses it rules out:

$$u_{\text{CG}}(a) := I_0(p(D)) - I_0(p(D|a)) \tag{5}$$

where $I_0(p(X))$ is a function that takes a discrete distribution $p(X)$ and counts over $x \in X$ how many values of $p(x) = 0$.

Causal Graphical Models

Equations 1–5 express four theories of how people assign values to the prospect of observing variables, given their probability beliefs $P(a)$, $P(d)$, $P(d|a)$. Previous work on information seeking has given these probabilities to participants, either by describing rates of events or by having participants infer those rates from experience—in both cases, the probabilities that get fed into their models are the ground truth probabilities.

But a cognitive theory should give an account of what probabilities people compute with (not necessarily the ground truth ones) and where these probability beliefs come from. People can reason on the basis of their theoretical knowledge of events whose probabilities they've never experienced firsthand (on the basis of e.g., causal knowledge that's been described to them). CGMs give a way to model such theoretical knowledge, representing causal knowledge as graphs, with variables as vertices and causal relationships as directed edges between vertices. CGMs have had success explaining behavior in a large range of tasks, and one question is whether CGMs, when used in conjunction with the right utility function(s), may underlie people's reasoning in an information-seeking decision task.

[1]Technically, D should appear as a free variable in Equation 1; we omit it for concision of notation.

A CGM generates a probability distribution over the values of its variables. It is common to assume that exogenous influences on a CGM's variables are uncorrelated, which entails (1) that CGMs satisfy the *causal Markov condition*, i.e., that in CGMs, each variable, conditioned on its parents, is independent of all its nondescendents; and (2) that we can express a model's joint distribution as

$$P(v_1, ..., v_N) = \prod_{j=1}^{N} P(v_j|\text{Parents}(v_j)) \tag{6}$$

where each v_j is a CGM variable. Assuming binary variables, that causes makes their effects more likely, and that causes integrate according to a noisy-OR function, each factor in that product can be computed as:

$$P(v_j = 1|\text{Parents}(v_j)) = 1 - (1 - b_j) \prod_{v_i \in \text{Parents}(v_j)} (1 - m_{ij})^{I_1(v_i)} \tag{7}$$

where b_j is the effect of background causes on variable v_j in the model, m_{ij} is the strength of the causal relation between v_j and parent v_i, and $I_1(v_i)$ is an indicator function that returns 1 when v_i is present, otherwise 0. If instead causes are assumed to integrate conjunctively (rather than via a noisy-OR function), then the probability of a variable conditioned on its parents is instead:

$$P(v_j = 1|\text{Parents}(v_j)) = 1 - (1 - b_j)(1 - m_j)^{I_1(\text{Parents}(v_j))} \tag{8}$$

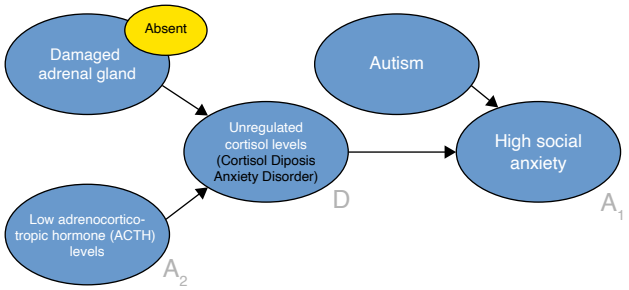
where b_j is the effect of background causes on variable v_j, $I_1(\text{Parents}(v_j))$ is an indicator function that's 1 when all of v_j's parents are present, else 0, and m_j is the causal strength of the conjunction of v_j's parents on v_j.

To see how CGM causal reasoning informs information seeking, take a situation in which a doctor is tasked with choosing to run a test for a disease and is deciding between two alternatives: A_1 and A_2. A general theory of information search would explain which test they choose to run (which observation they choose to make) in two phases. First, the doctor's internal model (representable as a CGM) generates distributions $P(D)$ (their prior beliefs about the disease's prevalence), $P(A_1)$ and $P(A_2)$ (their beliefs in the chances of test outcomes), and $P(D|A_1 = 0)$, $P(D|A_1 = 1)$, $P(D|A_2 = 0)$ and $P(D|A_2 = 1)$ (their beliefs in the conditional probability of the disease given those test outcomes).

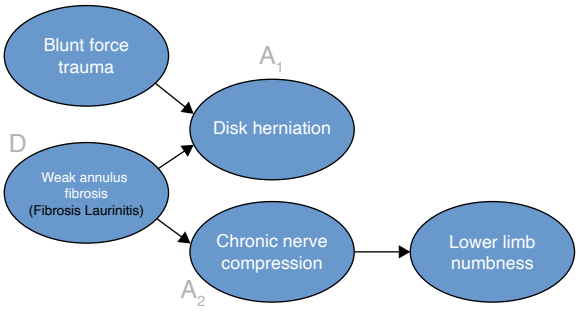
Then, the relevant distributions are fed into a utility function: $P(D)$, $P(A_1)$, $P(D|A_1 = 0)$, $P(D|A_1 = 1)$ as the inputs for A_1, $P(D)$, $P(A_2)$, $P(D|A_2 = 0)$, $P(D|A_2 = 1)$ as the inputs for A_2, and the utility function generates the informational value $U(A_1)$ of observing A_1 and $U(A_2)$ of observing A_2. The doctor chooses to run the test with the higher value.

The current study tested this theory of information search in a simulated medical context, where participants were given a disease's causal structure and tasked with choosing to run

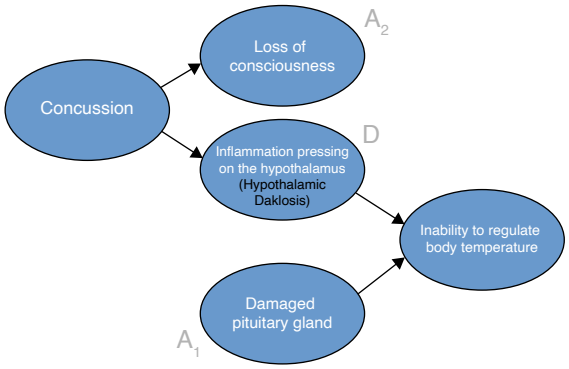
Cortisol Diposis Anxiety Disorder (CDAD). Patients with CDAD fail to regulate cortisol levels in their blood. Cortisol is produced and regulated by the hypothalamic-pituitary-adrenal system (HPA), and is secreted in response to stress. The two main causes of CDAD are damaged adrenal glands (the gland in charge of secreting cortisol) and low adrenocorticotropic hormone (ACTH- the hormone that causes the adrenal glands to secrete cortisol) levels. Both of these conditions lead to imbalanced cortisol levels thus, physical responses significantly differ from expected responses in social situations. This stress and anxiety is notably high in situations associated with a sudden change in plans, routine, mannerisms or movement. For that reason, CDAD is often confused with austism, a condition also known to cause significantly high levels of social anxiety.



Fibrosis Laurinitis (FL) is a condition in which the annulus fibrosis, the outer component that holds back disks together, is significantly weak. Patients that inherit Fibrosis Laurinitis suffer from multiple disk herniations and chronic nerve compression. The compression slowly (but constantly) wears down the integrity of the nerves around the spine, inevitably causing total lower limb numbness, typically by age sixty. Unfortunately, this condition cannot be cured, but the pain and numbing process can be reduced and slowed with heat, medication, and physical therapy.



Hypothalamic Daklosis (HD). Concussions associated with anterior damage lead to both a loss of consciousnessand rapid swelling, which presses on the hypothalamus, the part of the brain that regulates body temperature, among many other physiological functions. Increased pressure on the hypothalamus (a condition known as known as hypothalamic daklosis), therefore results in an inability to regulate body temperature. This highly dangerous condition creates a high risk of hypothermia or overheating. One must also be aware that damage to the pituitary gland also causes temperature deregulation, and is frequently considered when diagnosing hypothalamic daklosis.



Brasee-Fox Syndrome (BFS) is a condition in which the body's melatonin levels rapidly increase to potentially threatening levels. In patients with a damaged pineal gland (whether because of injury, infection, trauma, etc.) **and** intense ultraviolet light exposure (meaning **both** of these factors must be present), the body begins to significantly increase melatonin production. You will see this represented in the diagram below by two arrows converging to one arrow. Thus, a patient will only have Brasee-Fox syndrome if both of these causes are present. Patients who take too much supplemental melatonin, can however, also suffer from Brasee-Fox Syndrome. Overmedication leads to an sudden increase in melatonin levels, resulting in temporary symptoms. Patients that are overmedicated will also feel extremely lethargic.

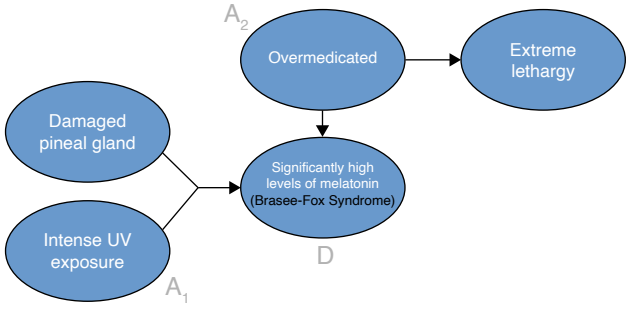


Figure 1: The four disease networks (reproductions). Oval nodes represent binary variables, arrows represent causal connections. The CDAD network had the "Damaged adrenal gland" absent across all questions. The gray labels are shorthand added for expositional convenience; participants did not see these. The BFS network is the only network with a description mentioning conjunctive causal combination.

one of two tests for the disease. In addition to asking participants to make information-seeking judgements, we also elicited the marginal and conditional probability judgements that are the constituents of the utility function described above. Doing so will allow us to distinguish if subjects' suboptimal information-seeking judgments are a result of (a) a faulty calculation of utility or (b) if utility is being calculated correctly but with faulty probability estimates.

Methods

Subjects

32 New York University undergraduates, aged 18–22, participated in this experiment. Participants were selected on a volunteer basis using SONA, an online scheduling program.

Procedure

Each participant learned and then answered questions about four diseases, presented one at a time. For each disease, a participant would be presented with a causal network representation of the disease, asked to make a forced-choice information-seeking decision, and asked for their probability judgements concerning disease network variables. HIV was used as a sample disease to ensure that participants understood the instructions. After answering five sample probability judgement questions and one sample forced choice question, the participant was given feedback, starting the actual task after understanding the answers to all sample questions.

Materials. Each participant learned about one disease for each network. Figure 1 shows the four disease networks. Participants read an introductory paragraph about each disease,

examined a causal network diagram for the disease and questions about the disease presented on a computer screen, and recorded their answers to the questions into their "diagnostic evaluation journal," a packet with a page per disease containing all the questions to be answered for that disease.

Disease presentation. In each of the four disease networks, one node signified the disease, the remaining nodes representing causes and symptoms related to the disease. Participants read a paragraph for each disease describing the causal relations in the diagrams (Figure 1). Henceforth, we'll refer to the networks by their initials: CDAD, FL, HD, and BFS. In the diagrams, edges between nodes signified causal relations between the variables, and participants were told that the presence of a cause entails the presence of its effect 100% of the time. Participants were told that all known information about each disease was included in the diagrams (i.e., that there were no unseen background causes). Participants were also told that during trials, a small yellow oval labeled "present" or "absent" overlayed over a blue variable oval would signify the variable's value if it was known.

Why these disease networks? Each disease network corresponds to a distinct reasoning task. The CDAD network involves an alternative cause structure—A_1 is perfectly diagnostic for D when autism is absent; if autism is present, then A_1's presence could be caused by either autism or D. By contrast, A_2 is always perfectly diagnostic of D. FL involves another alternative cause structure, but A_2 is now perfectly diagnostic of D by virtue of being D's effect. HD involves an indirect inference—A_2 is perfectly diagnostic of D, but requires one to reason from A_2 backward to whether the patient has a concussion, then forward to D. A_1's value is completely independent of D's. Finally, BFS involves a conjunctive cause (see Equation 8)—A_1 is only perfectly diagnostic of D when the a participant knows whether the pineal gland is damaged, whereas A_2 is always perfectly diagnostic.

Forced choice. During the forced choice task, each participant chose to learn the value of one of two variables. The two variables were designated with small yellow ovals containing three question marks. For example, for the FL network a participant would see the small yellow ovals overlaid on the "Disk herniation" and "Chronic nerve compression" ovals. The text inside the variable of interest (the D node), was made red ("Weak annulus fibrosis" for FL). The participant recorded in their journal which variable they'd choose to observe and rated their confidence on a seven-point scale. The order of the forced-choice questions and the set of probability judgements was randomized.

Probability judgements. Participants were also to produce judgements for the seven probabilities needed to generate utility values, as mentioned above: $P(D)$, $P(A_1)$, $P(D|A_1 = 0)$, $P(D|A_1 = 1)$, $P(A_2)$, $P(D|A_2 = 0)$, and $P(D|A_2 = 1)$. For example, for $P(D)$ in the HD network they were be asked, "Does [patient initials] have inflammation pressing on the hypothalamus (Hypothalamic Daklosis)?" and they would respond "YES" or "NO" and give their confidence (0%–100%). For $P(D|A_1 = 0)$ in the BSF network, a participant would be shown the BSF network with a small yellow "Absent"-labeled oval overlaying the "Intense UV exposure" oval and asked, "Does [patient initials] have "Significantly high levels of melatonin (Brasee-Fox Syndrome)?" The order of the probability judgement questions was randomized across diseases and participants. Though we stipulate in the instructions that causal links are deterministic, we nonetheless ask for probability judgements on the grounds that (1) not all participants may attended to these instructions and that (2) some subjects may have used their real-world knowledge to conclude that the causal relations were probabilistic despite the instructions.

Modeling

The forced-choice and their corresponding confidence ratings were combined into a confidence-adjusted choice score (CACS), ranging between 0 (complete confidence in A_1) and 1 (complete confidence in A_2). 0.5 represents no preference for either alternative. Each of a participant's models aims to predict their four CACSs, making assumptions about how they compute with probabilities and where those probabilities come from. For each disease network, all models use the seven probabilities $P(D)$, $P(A1)$, $P(A2)$, $P(D|A1 = 0)$, $P(D|A1 = 1)$, $P(D|A2 = 0)$, $P(D|A2 = 1)$ to compute $U(A1)$ and $U(A2)$, then crucially use the difference $U(A2)$ - $U(A1)$ to predict the the participant's CACS for that network using Eqs. 2-5.

We used two broad kinds of models to produce CACS predictions, *direct models* and *causal models*. The direct models fed a participant's probability judgements directly into a utility function. There were four direct models per participant, one for each of the four utility functions. The causal models instead used CGMs to generate probability judgements, and fed those probability judgements to utility functions. Each causal model had three parameters: the background cause parameter for the root nodes (b_r), the background cause of the non-root nodes (b_n), and a single causal strength parameter (m). The causal models can be thought of as adding a constraint on the relationships between a participant's probability beliefs, namely, the constraint that a participant's probability beliefs be consistent with a CGM. We're interested, then, in understanding whether adding this constraint will improve fits to participant choice scores in the information-seeking task, i.e., in understanding whether people use CGM-based causal knowledge when searching for information.

We created three kinds of causal models: deterministic and quasi-deterministic causal models, and fit causal models. The *deterministic causal model* had ($b_r = 0.5$, $b_n = 0$, $m = 1$)—non-root nodes were never present in the absence of all of their parents and never absent in the presence of at least one parent. These can be considered "ground truth" models, just in the sense that they're prima facie consistent with the task instructions (no background causes, deterministic causal links). The *quasi-deterministic causal model* had ($b_r = 0.5$, $b_n = 0.1$, $m = 0.95$)—allowing for non-root nodes were occa-

sionally though rarely present in the absence of their parents and occasionally though rarely absent in the presence of their parents. There were 4 deterministic models and 4 probabilistic models, for the same reason there were 4 direct models.

Finally, the third kind of causal model—the *fit causal models*—were fit to participant probability judgements by fitting (b_r, b_n, m) to minimize the squared-error between the CGM-generated and actual probability judgements; 4 utility functions gives a total of 4 fit causal models. Note that though each of the causal models predicts four CACS scores per participant (corresponding to their choices on the four networks), each causal model holds fixed its parameters across the disease networks—we assume that a participant has a single generic set of causal parameters that they use for all four networks.

In total, we have 16 models (4 direct, 4 det., 4 quasi-det., 4 fit) per participant.

Scaling. Each of the 16 models uses probabilities to compute $U(A2) - U(A1)$, a predictor of CACS. As $U(A2) - U(A1)$ is not constrained to be in $[0,1]$, we allowed predictions to deviate via a linear transformation from the CACSs they were predicting—for each of the 16 models, we fit a slope and an intercept term to minimize the squared error between a participants four CACSs and the four $U(A2)-U(A1)$ predictions generated from their models.

Results

Generally, the mean CACSs across participants were larger than 0.5, indicating that participants tended to prefer A_2 to A_1: the mean CACS was 0.72 for the CDAD network ($t(31) = 3.74$), 0.82 for the FL network ($t(31) = 7.52$), 0.76 for the HD network ($t(31) = 5.07$), and 0.80 for the BFS network ($t(31) = 7.37$). $p < .001$ for each of comparisons. For each of the networks, A_2 is the "normative" response—see the "Why these networks?" subsection in the previous section, so that we can say that participants generally seem to be reasonable in their choices.

As we have 16 models per participant, we used Bayesian model averaging to derive predictors that incorporated our uncertainty over models. First, for each of a participant's 16 models we derived a BIC from the squared error between the model's predicted CACS and the participant's actual CACS:

$$\text{BIC} := n \log\left(\frac{\text{SSE}}{n}\right) + K \log(n) \tag{9}$$

Each of a participant's models generates four CACSs (one per network), and so $n = 4$ always. For the direct models, deterministic models, and quasi-deterministic models $K = 2$ (the two scaling parameters). For the fit models, $K = 2$; K doesn't increase for the models because we fit our models not to the CACSs (which were used to compute BIC-determining SSE) but rather to a participant's probability judgements.

From these BICs, following Neath and Cavanaugh (2012), we derived for each participant a posterior distribution over their 16 models. As each model makes four CACS predictions and is associated with a posterior probability, we can compute the expected value of the four CACS predictions, taking the expectation over models. Averaged across subjects, the model predictions were 0.707, 0.815, 0.779, and 0.795 for the CDAD, FL, HD, and BFS networks. Figure 2 shows the predicted (Bayesian-model averaged) vs. actual CACSs across participants. Fit quality is good. We can now use the model posteriors to compute marginal distributions in order to understand what utility function participants seem to favor and whether or not it's worth using CGMs to generate predictions.

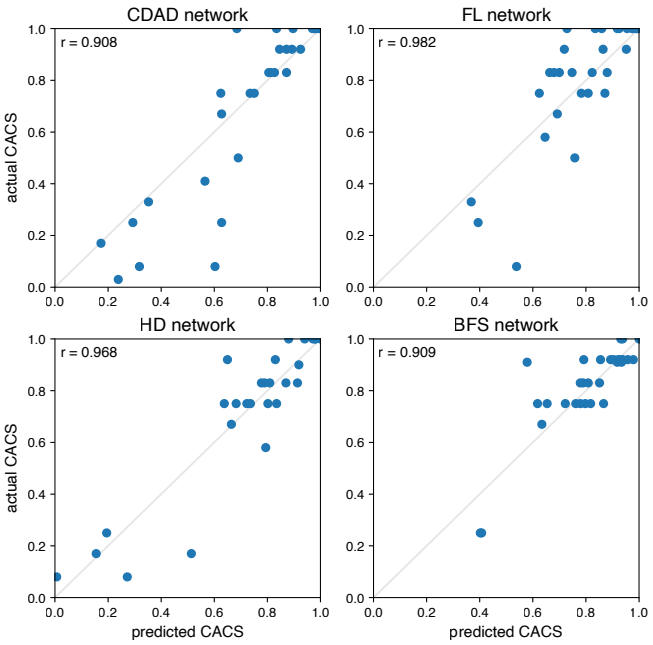


Figure 2: Predicted vs. actual CACSs, plotted across participants. All correlations are significant ($p < .001$).

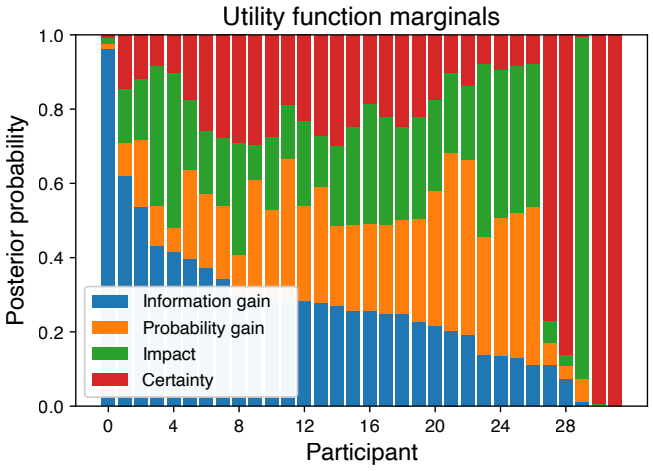


Figure 3: Utility function posterior distributions across participants.

Marginalizing over all but the utility function, we find that $p(u_{IG}) = 0.271$, $p(u_{PG}) = 0.222$, $p(u_{Imp}) = 0.241$, and $p(u_{CG}) = 0.266$. Figure 3 shows the variation across participants in the distribution over utility functions. Across 32 participants, information gain is the MAP utility function for 9, probability gain for 7, impact for 11, and certainty for 5. There seems to be substantial individual variation in the posterior distributions, enough variation to make it difficult to conclude a clear winner.

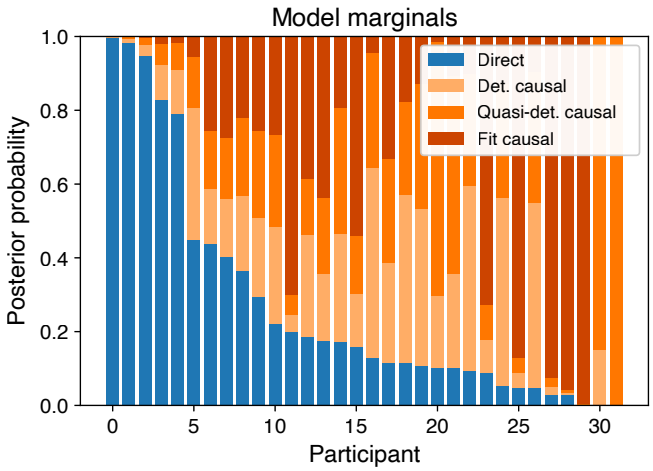


Figure 4: Model posterior distributions across participants.

Marginalizing over all but model kinds, we find that $p(\text{direct}) = 0.271$, $p(\text{det.}) = 0.202$, $p(\text{quasi-det.}) = 0.237$, $p(\text{fit}) = 0.289$. Figure 4 shows the variation across participants in the posterior distribution over models. Across 32 participants, the MAP model is the direct model for 10, the deterministic model for 6, the quasi-deterministic model for 5, and the fit causal model for 11. Notably, it's worth it for us to constrain the probability judgements of the majority (22/32) of participants with some CGM or other.

Discussion

Information search is an important ability both in everyday and specialized contexts, contexts in which causal knowledge is ubiquitous; this study is, to our knowledge, the first examining information search in an explicitly causal setting. Generally, participant choices reveal they tend to seek information rationally. We presented participants with four causal information-seeking scenarios, each requiring a different kind of causal reasoning, and found that responses during a choice task tended to be consistent with probabilities generated by a CGM.

In fact, for two-thirds of participants, the best (MAP) model was a model constraining probabilities to be consistent with some CGM or other. On the basis of nothing but descriptions of unfamiliar events in a hypothetical causal scenario, participants were able to make use of the provided causal knowledge to seek information—the probability judgements that best explained their behaviors were ones that were constrained to be consistent with a formalization of causal knowledge.

There is a good amount of variation across participants in preferred utility function, though it's difficult to tell whether this variation reflects the distribution of people's preferred utility functions, or whether it reflects a limitation in our design: The four causal scenarios we tested were not chosen to distinguish between the different utility functions.

We could, in an extension to this study, stipulate nondeterministic causal strengths ($m < 1$) and possible background causes ($b_n > 1$). A cover story might involve machines that cause each other to take a state with some probability, from which state they'd have some chance of causing other machines to take a state, etc. Aside from enabling us to ask how people reason in nondeterministic scenarios, this extension could allow us to design "more optimal" experiments (Coenen, Nelson, & Gureckis, 2019), where we would, following Nelson et al. (2010), search for causal model parameters that maximally distinguish between the CACSs anticipated by different utility functions.

References

Baron, J., Beattie, J., & Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning ii. congruence, information, and certainty. *Organizational Behavior and Human Decision Processes*, *42*, 88–110.

Baron, J., & Hershey, J. C. (1988). Heuristics and biases in diagnostic reasoning: I. priors, error costs, and test accuracy. *Organizational Behavior and Human Decision Processes*, *41*, 259–279.

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

Coenen, A., Nelson, J. D., & Gureckis, T. M. (2019). Asking the right questions about the psychology of human inquiry: Nine open challenges. *Psychonomic Bulletin & Review*, *26*, 1548–1587.

Eddy, D. M. (1982). Probabilistic reasoning in clinical medicine: Problems and opportunities. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 249–267). Cambridge, MA: Cambridge University Press.

Glymour, C. (1998). Learning causes: Psychological explanations of causal explanation. *Minds and Machines*, *8*, 39–60.

Griffiths, T. L., & Tenenbaum, J. B. (2005). Structure and strength in causal induction. *Cognitive Psychology*, *51*, 334–384.

Hagmayer, Y., & Sloman, S. A. (2009). Decision makers conceive of themselves as interveners. *Journal of Experimental Psychology: General*, *128*, 22–38.

Holyoak, K. J., Lee, J. S., & Lu, H. (2010). Analogical and category-based inferences: A theoretical integration with bayesian causal models. *Journal of Experimental Psychology: General*, *139*, 702–727.

Kemp, C., Shafto, P., & Tenenbaum, J. B. (2012). An integrated account of generalization across objects and features. *Cognitive Psychology*, *64*, 35–73.

Klayman, J., & Ha, Y. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*, 211–228.

Lindley, D. V. (1956). On a measure of the information provided by an experiment. *The Annals of Mathematical Statistics*, *27*(4), 986–1005.

Markant, D. B., & Gureckis, T. M. (2014). Is it better to select or to receive? learning via active and passive hypothesis testing. *Journal of Experimental Psychology: General*, *143*(1), 94.

Neath, A. A., & Cavanaugh, J. E. (2012). The bayesian information criterion: background, derivation, and applications. *Wiley Interdisciplinary Reviews: Computational Statistics*, *4*(2), 199–203.

Nelson, J. D., McKenzie, C. R., Cottrell, G. W., & Sejnowski, T. J. (2010). Experience matters: Information acquisition optimizes probability gain. *Psychological Science*, *21*(7), 960–969.

Oaksford, M., & Chater, N. (1994). A rational analysis of the selection task as optimal data selection. *Psychological Review*, *101*(4), 608.

Pearl, J. (2000). *Causality: models, reasoning, and inference*. Cambridge, UK: Cambridge University Press.

Rehder, B. (2003). A causal-model theory of conceptual representation and categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *29*, 1141–1159.

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

Rehder, B., & Hoffman, A. (2005). Eyetracking and selective attention in category learning. *Cognitive Psychology*, *51*, 1–41.

Rehder, B., & Kim, S. (2010). Causal status and coherence in causal-based categorization. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *36*, 1171–1206.

Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*, *20*, 273–281.