Unifying inference and selection in singular causal explanation

Anonymous CogSci submission

Abstract

Explaining why events occurred involves solving different information-processing problems: inference about what actually happened (causal inference) but also highlighting a subset of the causes that contributed to the outcome (causal selection). Although past research has investigated causal inference and causal selection separately, we report results of an experiment (N=284) examining how people solve both problems jointly, as is the case in real-world explanation settings. We find evidence that participants infer the state of unobserved variables on the basis of available evidence, and observe common behavioral signatures of causal selection. However, explanation preferences deviate in important ways from the predictions of a computational model combining existing theories of causal inference and causal selection. In particular, participants were disproportionately likely to cite unobserved variables. We suggest a possible preference for producing explanations that allow the explainee to benefit from inferential work performed by the explainer.

Keywords: causality; counterfactuals; explanation; inference

Introduction

Why did this car accident happen? Why did the dinosaurs become extinct? The drive to explain why a particular event happened is one of the core psychological features of our species, and a common topic of discussion and debate. In the field of causal cognition, this problem of *singular causal explanation* has received a large amount of attention (Lombrozo, 2006; Woodward, 2021; Lagnado, 2021). Providing a causal explanation typically involves solving several different information-processing problems. In this paper we focus on two of the most important:

-**Causal inference** is the problem of using one's causal beliefs to figure out what happened on the basis of the available evidence. For example, given that the driver was coming back from a party, how likely is it that he was drunk? Given that it was very cold that night, how likely was it that there was ice on the road?

-**Causal selection** refers to the problem of highlighting one cause out of the several causes that contributed to an outcome (Hesslow, 1988; Quillien & Lucas, 2023). Suppose we know the driver was drunk, that there was ice on the road, and that both factors contributed to the accident. Which fact should we highlight as *the* cause of the accident?

It is easy to see that solving both problems is crucial to successful causal explanation in everyday cases. The details of what happened are rarely all transparently observable, so someone looking for an explanation typically needs to piece them together from the available evidence. And in the real world any given outcome is the end result of a complex net of many causes, so selection is necessary to avoid producing overwhelmingly complex explanations.

In the existing literature, these problems have almost exclusively been studied separately from each other. In this paper we study how people give causal explanations when they have to jointly solve both problems. We sketch a computational framework for causal explanation in the presence of unobserved variables, and we report the results of an experiment testing the predictions of this model.

Background

Inference and selection in singular causal reasoning

A large literature has explored how people make inferences about whether an event happened, on the basis of information about the other events that happened. In a setting where events are causally related to each other, this is a problem of causal inference, and it can be solved using the normative formalism of causal graphical models (Pearl, 2009). Many experiments have found that people make inferences in ways that approximate the normative prescriptions of causal models (Sloman & Lagnado, 2004; Hagmayer et al., 2007; Lagnado, 2021), although with noteworthy deviations (Davis & Rehder, 2020).

In our experiment we will focus on diagnostic inferences: inferring the value of a potential cause, after observing the effect (as well as other potential causes) occurring (see Pearl (2009); Meder & Mayrhofer (2017) for a detailed treatment of inference on causal networks). To give a very simple example, suppose that event C often causes event E, we observe that E happens, and we want to infer the probability that C happened. We can solve this problem using Bayes' rule:

$$P(C|E) = \frac{P(E|C)P(C)}{P(E)} \tag{1}$$

where the likelihood $P(E|C)$ depends on the parameters of the causal model describing the causal system.

In contrast, research on causal selection investigates how reasoners judge which of the factors that contributed to an outcome is the most important cause (Hesslow, 1988). For example, although the presence of the oxygen in the air and a bolt of lightning both contributed to a forest fire, most people have the intuition that the lightning is *the* cause of the fire. For simplicity, extant research on causal selection has used experimental settings where the reasoner already knows what happened. Because of this, not much is known about causal selection in contexts where people also need to make inferences about what happened.

In this paper, we study causal explanation in a context where some events are unobserved. For example, participants are told that a reading group is successful if both the professor and the postdoc attend, provided that they both talk about the assigned paper. Participants know that today, both the professor and the postdoc attended, and the reading group was successful; but they are not given any information about who talked about what.[1] Why do participants think the reading group was successful? This task requires causal selection (because there are four potential causes) as well as inference (because of the unobserved events). In the next section we outline a computational framework for causal explanation in this setting.

Computational framework

Throughout we assume that the reasoner knows the causal structure of the relevant system, and we make use of the formalism of Structural Causal Models, in which *variables* represent whether a given event occurs (for example $C = 1$ means that event C happened), and *structural equations* determine the causal relationships between variables (see Pearl (2009) for details).

We consider a causal system where two variables A and B can have a causal influence on outcome variable E. For each cause variable X there is an associated unobserved variable X_u that determines whether X can have an effect on E. Figure 1 shows a graphical model of such a causal system. We study a disjunctive and a conjunctive structure. In the disjunctive structure E happens if either both A and A_u happen or both B and B_u happen:

$$E := (A \wedge A_u) \vee (B \wedge B_u) \tag{2}$$

In the conjunctive structure E happens if all other variables happen:

$$E := (A \wedge A_u) \wedge (B \wedge B_u) \tag{3}$$

While the values of A and B are observed, the values of A_u and B_u are not. To give a causal explanation for why E happened, the reasoner must i) infer the value of A_u and B_u, ii) engage in causal selection, iii) integrate the two processes. We discuss each component in turn.

Causal inference

We assume that the reasoner infers the values of A_u and B_u by using Bayes' rule:

[1] Or, in fact, who has even done the reading. But that's another story.

$$P(A_u, B_u | A, B, E) = \frac{P(E|A_u, B_u, A, B)P(A_u, B_u)}{P(E|A, B)} \tag{4}$$

Causal selection

According to an increasingly popular family of accounts, people engage in causal selection by imagining counterfactual possibilities (Icard et al., 2017; Quillien, 2020; Henne et al., 2019), see also Gerstenberg et al. (2021). Here we use a recent computational model of causal selection based on this idea.

The *Counterfactual Effect Size Model* (CES model Quillien, 2020; Quillien & Lucas, 2023) holds that people judge whether event C was a cause of event E by i) simulating many different alternative ways the situation could have happened ii) computing a measure of the dependence between C and E across these possibilities.

Each counterfactual possibility is simulated by sampling each cause variable from a probability distribution, and then setting the effect variables according to their structural equations. Each cause variable V is sampled from the probability distribution $s\delta(V) + (1-s)P(V)$, where $\delta(V)$ is the value of V in the actual world, $P(V)$ is the prior probability of V, and s is a 'stability' parameter (we will set $s = .7$ on the basis of past empirical data, Lucas & Kemp (2015); Quillien & Lucas (2023)).

The CES score of C for E is then computed on the basis of the simulated possibilites. In our setting, it is equivalent to the Pearson correlation coefficient between C and E across the simulated counterfactual possibilities.

The CES model has successfully explained data from past experiments on causal judgments (Lagnado et al., 2013; Gerstenberg & Icard, 2020; Icard et al., 2017; Morris et al., 2019) and made successful new predictions, both in simple experimental settings (Quillien & Lucas, 2023; Konuk et al., 2023) and in a real-world context (Quillien & Barlev, 2022). However, to our knowledge it has not been tested in settings like ours where the state of some variables is unobserved.

Causal explanation with unobserved variables

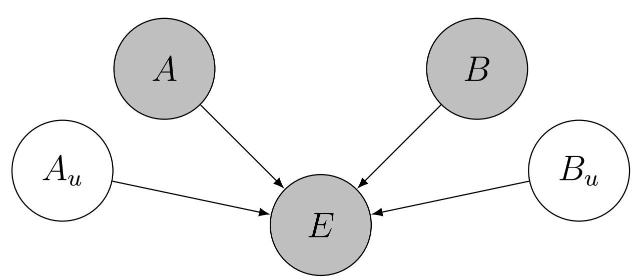

Figure 1: Directed Acyclic Graph depicting the causal structure in our experiment. Grey nodes denote observed variables, white nodes denote unobserved variables.

We now offer a model of causal judgment that integrates the two components above. Our goal is to assign an overall causal score $C(X=x)$ to each event, such that an event with a higher causal score is a better candidate for a causal explanation. For convenience we will denote the posterior distribution in abbreviated form as $P(A_u, B_u|A, B, E) = P_\alpha(A_u, B_u)$. It will also be useful to define $K(X=x)$ as the CES score of event $X=x$. We will also write $K(X=x|\mathbf{V}{=}\mathbf{v})$ to express the CES score that would be assigned to $X=x$ under the assumption that **V=v** in the actual world (this is useful notation when we need to consider several possible hypotheses about the actual world consistent with our observations).

The CES model is defined for situations where we already know the full state of the world. To apply it to the present case (where this assumption doesn't hold), we must make some choices as to how to handle the uncertainty over A_u and B_u.

One intuitive way to do this is to compute a CES score for each possible state of the actual world compatible with what we know, and then compute a weighted average of these scores, where the weights are the probabilities of the states of the world. For example to compute the CES score for $A=a$, denoted $K(A=a)$, we compute:

$$K(A=a) = \sum_{A_u, B_u} K(A=a|A=a, B=b, A_u, B_u) P_\alpha(A_u, B_u) \tag{5}$$

where a and b are the actual-world values of A and B, and $K(X=x|\mathbf{V}=\mathbf{v})$ is the CES score we would compute for $X=x$ if we knew that the actual-world values of $\mathbf{V}$ were $\mathbf{v}$.

Computing the CES score for the unobserved variables introduces one additional complication: we typically don't know whether the variable has value 1 or 0. One intuition is that people will tend to say '$A_u = 1$ caused the outcome' if i) it is in fact likely that $A_u = 1$ in the actual world, ii) $A_u = 1$ has a high CES score. One way to implement this is to compute C by multiplying the CES score K by the posterior probability of the variable value. For example for $A_u = 1$:

$$\begin{aligned} C(A_u=1) &= K(A_u=1)P_\alpha(A_u=1) && (6) \\ &= \sum_{B_u} K(A_u=1|A=a, B=b, A_u=1, B_u) \\ &\quad \times P_\alpha(B_u|A_u=1)P_\alpha(A_u=1) && (7) \\ &= \sum_{B_u} K(A_u=1|A=a, B=b, A_u=1, B_u) \\ &\quad \times P_\alpha(A_u=1, B_u) && (8) \end{aligned}$$

Actual causation

In addition to causal selection, people also engage in a more categorical kind of causal judgment, differentiating between variables that had at least some contribution to an outcome and those that did not contribute at all. In our computational model we use a simple heuristic to exclude events that do not qualify as actual causes. Specifically, we assign a causal score of $C(X) = 0$ to any variable X whose value does not match the value of the outcome (e.g., if $E = 1$, then $B = 0$ is not an actual cause of E) and to unobserved variables if their observed counterpart has value 0. For more sophisticated computational accounts of categorical actual causation see for example Halpern (2016).

General mathematical framework

Here we give a more general formalization of our proposal, generalizing from the examples above. We consider whether variable realization $X=x$ was the cause of outcome $E=e$. We denote $\mathbf{V}$ the set of variables other than E and X. The CES score K of $X=x$ is computed by i) assuming that $X=x$ in the actual world, and ii) marginalizing across all possible values of the other variables, weighted by their posterior probabilities:

$$K(X=x) = \sum_{v \in V} K(X=x|\mathbf{V}=\mathbf{v}, X=x) P_\alpha(\mathbf{v}|X=x) \tag{9}$$

The overall causal score $C(X=x)$ is then computed by weighing K by the posterior probability of $X=x$. We also check for actual causation. Formally:

$$C(X=x) = K(X=x)P_\alpha(X=x)T(X=x) \tag{10}$$

where $T(X=x)$ is 1 if $X=x$ is an actual cause of E, and 0 otherwise.

Softmax choice model

The sections above specify how the model assigns causal scores to variables. To convert these causal scores to predicted choice proportions, we assume that participants are soft-maxing over the causal scores:

$$P(\text{choice} = X) \propto \exp(\frac{C(X)}{\tau}) \tag{11}$$

where τ is a temperature parameter (higher values indicate more stochasticity) that we fit to the data.

Lesioned models

We will also explore 'lesioned' models to assess our claim that when people make a causal judgment, they engage both in inference (about the value of unobserved causes) and in causal selection.

Lesioning inference Our first lesioned model lesions the inference module. That is, we have $P_\alpha(A_u, B_u) = P(A_u, B_u)$. In words, instead of setting $P_\alpha(A_u, B_u)$ to be the posterior, we 'freeze' it as the prior distribution. The model otherwise works exactly as above.

Lesioning causal selection The second model lesions the causal selection module. We assume that people do not engage in counterfactual simulation when making causal judgments. Once they determine which variables are actual causes of E, they select these variables simply in function of their posterior probabilities. In terms of the mathematical framework defined above, we replace all CES scores K by 1.

Lesioning both inference and selection This model assumes that people select among actual causes almost indiscriminately. That is, they assign a causal score of $C = 1$ to observed variable values, and a causal score of $C = P(X_u = x_u)$ to unobserved variables, where $P(X_u = x_u)$ is the prior probability of that variable.

Lesioning actual causation We will also test variants of the models defined above that do not check if an event is an actual cause of the outcome.

Methods

We conducted a behavioral experiment to test our models. You can see it in our Anonymised Repository.

Design

We asked each participant to give causal explanations for variable E's occurrence or non occurrence across the 12 different logically possible combinations of observed variables ('worlds'), five of which used the conjunctive structure defined in Equation 3, and seven used the disjunctive structure defined in Equation 2[2]. The underlying causal structure was presented as a simple story along with a simplified Directed Acyclic Graph. For example, in the 'reading group' story participants read:

> The situation is a small university reading group of students and their advisors. The students always attend, but the lecturer and postdoc only sometimes attend. Even when they attend, they do not always talk about the allotted paper. The lecturer attends [10%] of the time and the postdoc attends [80%] of the time. The lecturer talks about the paper [70%] of the times they attend, and the postdoc talks about the paper [50%] of the times they attend. A good discussion happens when [either] the lecturer or postdoc attend and talk about the allocated paper [...] On this occasion, the lecturer [attended] and the postdoc [did not attend] and there was [a] good discussion.

Participants were also given the prior probabilities of events, which we manipulated (see Table 1). On each trial they were shown an event (they knew the state of observed variables, but not of the unobserved variables), and asked to explain the outcome by selecting one explanation (see Figure 2).

The structure of the causal system was presented verbally as a vignette. We used three cover stories: 1) a cookery tv show (loosely based on Zultan et al. (2012)), 2) a university reading group and 3) a job interview. For each trial, one probability set and one cover story ('cookery show', 'reading group', or 'job interview') was randomly selected. Our analyses collapse across cover stories.

[2]The unequal split is due to the fact that some events are possible for the disjunctive structure but not the conjunctive structure (e.g., $A = 1, B = 0, E = 1$).

Table 1: Event probability manipulation: Three settings

Var	$P(\text{Var} = 1)$	Set 1	Set 2	Set 3
A		.1	.5	.1
Au		.5	.1	.7
B		.8	.5	.8
Bu		.5	.8	.5

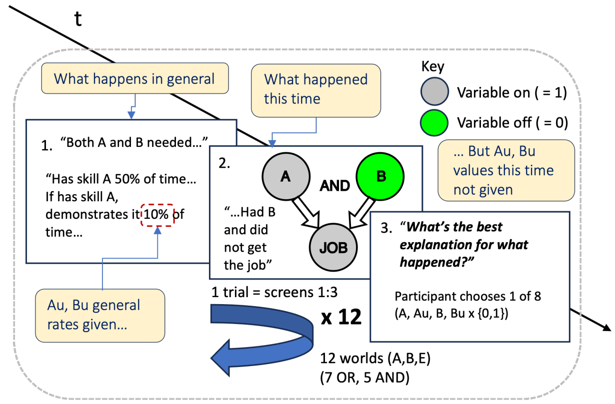

Figure 2: Simplified schematic of flow through of one trial

Participants

We recruited 284 fluent-English participants (125 female, 1 other, age Mean $\pm$ sd 36.8 $\pm$ 12.4, range 18-78) using the Prolific subject pool. They were paid £2.50 and the experiment took Mean $\pm$ sd 17.7 $\pm$ 7.8 minutes.

Stimuli

Each trial was a series of text and pictures following the same format (see Figure 2 for flow), created using JSPsych 6.3.1 html plugins (De Leeuw, 2015). The general schema presented the base rates at which the events usually happen, and the causal setup of the world (i.e. whether conjunctive — both events needed for the outcome to occur, or disjunctive — just one), and then described what happened to the observed variables *this time*.

Procedure

The experiment was implemented in JavaScript, hosted on Prolific and participants completed it in the browser on their own devices. After calibrating their computer screen, they were presented with the study's information sheet and consent form. Participants were then given instructions for completing the experiment and shown examples of the stimuli. They then completed a four-item quiz to test their understanding before beginning the experiment. All participants saw all 12 worlds one by one in a random order. The left/right presentation position on screen of the variables and their prior probabilities was counterbalanced between participants.

On each trial, participants were asked to explain the outcome by selecting one of eight possible explanations, for example: "The lecturer..." $\{$1. "did not attend" ($A = 0$); 2. "attended" ($A = 1$); 3. "...did not talk about the paper" ($A_u = 0$); 4. "...talked about the paper}" ($A_u = 1$)} and similarly for the postdoc {5. $B = 0$; 6. $B = 1$, 7. $B_u = 0$, 8. $B_u = 1$}.

Analysis

Data were analysed using R version 4.1. Package *lme4* (Bates et al., 2014) was used for mixed effects regression models following recommendations of Meteyard & Davies (2020), via package *lmerTest* (Kuznetsova et al., 2017) for tests. The Data and the R code for modeling and analysis are available in our Anonymised Repository.

Results

Figure 3 shows the choice proportions of participants and our full computational model. Firstly, people's judgments are clearly systematic (item-level goodness-of-fit χ^2 = 8874, df = 287, $p < .001^{***}$). Secondly, they choose actual causes over non-actual (on 88.2% of trials, χ^2 = 1993, $p < .001^{***}$); non-actual are the gray error bars in Figure 3. Thirdly, they choose unobserved variables over observed 61.2% of the time (χ^2 = 170.4, $p < .001^{***}$, see Figure 4 and subsection below).

Abnormal inflation as evidence for causal selection

Of special interest is whether participants' judgments exhibit the signature patterns of causal selection documented in past work (Morris et al., 2019). Human causal selection typically exhibits an effect called *abnormal inflation*, whereby people attribute greater causal responsibility to causes that are rare, infrequent or otherwise abnormal (Gerstenberg & Icard, 2020; Icard et al., 2017). In this analysis we focus on trials where $A = 1$, $B = 1$, $E = 1$ in both the conjunctive and disjunctive structure, because these are the trials that are closer to those investigated in past work on causal selection. In these trials, the CES model predicts abnormal inflation.[3]

To formally test whether participants reliably choose the abnormal variable (among the two observed variables A and B and excluding probability setting 2 from this analysis because A and B have the same probability), we ran a logistic mixed-effect regression predicting selection of the abnormal observed variable with random intercepts for condition and participant, on a dataset restricted as above. This shows a significant difference in the expected direction (odds ratio, estimate = .396, se = .349, CI [.204 .768], Z = -2.74, $p < .01^{**}$).

This result suggests that our experiment engaged some of the same cognitive mechanisms as other causal selection tasks. Since the abnormal inflation effect is predicted by a counterfactual model, the effect provides some evidence that this process of causal selection involved counterfactual reasoning.

[3]Note that the prediction for the disjunctive case contrasts with previous research which has found *abnormal deflation* (a preference for the most normal variable), in disjunctive structures (Gerstenberg & Icard, 2020; Icard et al., 2017). However, our disjunctive structure is slightly more complex than in this previous research, and consists in a disjunction of conjunctions (see Eq. 2). The CES model predicts abnormal inflation in this structure.

Unobserved vs observed variables

Participants could select an observed or unobserved event in their explanation. For example, they can say that the reading group was successful because the lecturer attended (an observed event), or because (presumably) the lecturer talked about the paper (an unobserved event that can be inferred from the available evidence). We find that participants i) preferred to select unobserved relative to observed events on average, ii) selected unobserved events to a larger extent than predicted by our main computational model, see Figure 4.

To test this effect, we sampled an explanation from the model for each participant observation. We ran a binomial logistic mixed-effect regression predicting 'answer unobserved' with a fixed effect for group (participant v model), and random effects for condition and participant. We found a main effect of group (odds ratios, estimate = 1.50, se = .052, CI [1.34 1.64], Z = 7.58, $p < .001^{***}$), whereby unobserved variables were cited more often by participants than by the model. We discuss this finding in the General Discussion.

Model fit

We fit the models to the full data from all conditions by minimising negative log likelihood, with the softmax temperature parameter τ as a free parameter, optimised with Brent method as implemented by R's `optim` function. See Table 2 for the model fits. The full model (containing the three modules of causal selection, inference and actual causation) fit well, but was beaten by the model lesioned to have no causal selection. The item-level Pearson correlation coefficient between the full model and participants' average judgments was $r(286) = .74$, $p < .001^{***}$, and between the best-fitting causal-selection-lesioned model and participants' average judgments was $r(286) = .78$, $p < .001^{***}$.

Model	τ	LogL	BIC
full	.299	-4680	9369
noActual	.309	-4646	9299
noInference	.208	-4883	9776
noSelection	**.361**	**-4304**	**8615**
noActnoInf	.490	-5487	10981
noActnoSelect	.342	-4675	9359
noInfnoSelect	.534	-5150	10308
noActnoInfnoSelect	.641	-5570	11147

Table 2: Temperature parameter τ and model performance metrics LogL and BIC.

The fact that lesioning the causal selection module improves the fit of the model is surprising given the presence of abnormal inflation in participants' judgments, an effect predicted by our causal selection model. This poor performance can be explained by the fact that the causal selection module tends to assign high causal responsibility to observed variables in situations where participants actually prefer unobserved variables. It also makes wrong predictions in many cases where the outcome does not happen ($E = 0$). In sum,

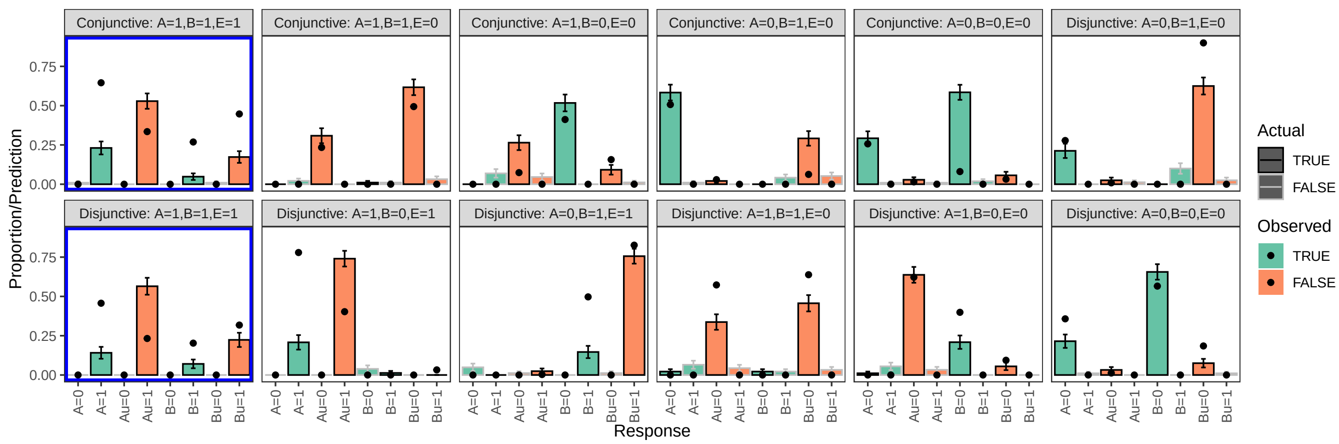

Figure 3: Results in Setting 3 $P(A) = .1, P(A_u) = .7, P(B) = .8, P(B_u) = .5$. Participants (bars) plus *Full* model (black circles). Blue highlights for canonical "everything happened" world ($A = 1$, $B = 1$, $E = 1$), expanded in Figure 4. See Anonymised Repository for plots of the other probability settings.

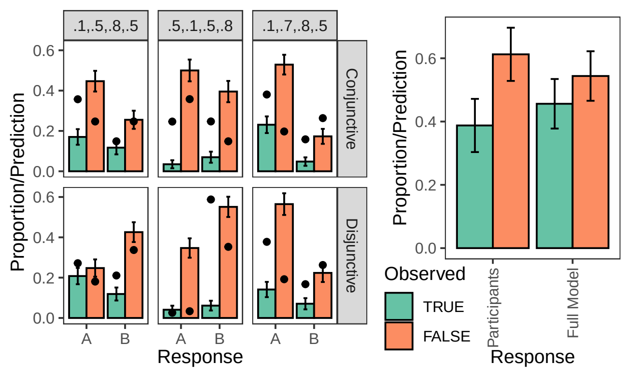

Figure 4: Left: Comparing $A = 1$, $B = 1$, $E = 1$ scenarios across probability settings 1-3 and conjunctive vs disjunctive structures. Upper facet labels show the probability of A, A_u, B and B_u in that order. Settings 1 and 3 show abnormal inflation in both structures for observed variables. Right: Overall propensity to select observed vs unobserved variables ($M \pm SE$ across worlds). Participants cite a larger proportion of unobserved variables than the full model.

while we have some evidence that participants are engaged in causal selection (they are not simply selecting randomly among observed causes of the outcome), our model does not fully capture how they do so.

In contrast, lesioning the causal inference module resulted in a worse fit (see Table 2). This result suggests that participants make approximately sound inferences about the probability that an unobserved event happened, and leveraged these inferences in their causal explanations.

Discussion

Causal explanation is a complex cognitive activity that requires solving multiple sub-problems. Research on causal cognition has typically focused on one sub-problem at a time: for instance some experiments focus on causal inference, while other experiments focus on causal selection. This strategy has been fruitful, but has also led to a neglect of the study of general problem of causal explanation where both problems are in play, as is typically the case in the real world. Here we considered how reasoners give causal explanations in a setting where some events are unobserved, such that reasoners need to engage in both causal inference and causal selection. First, we sketched a computational framework for how these two processes might be integrated by the mind. Then we reported the results of an experiment testing how people give causal explanations in this setting.

Our experimental data suggests that people engage in inference and selection in a way that is partially predicted by existing theories of these processes. At the same time, we also uncover phenomena that are not predicted by our computational framework based on this work. In particular, we find that people prefer to explain an outcome by citing an unobserved event, rather than an observed event, and that preference is stronger than predicted by our model.

We speculate that this finding reflects the fact that the explainer had to perform some computational work to infer whether the unobserved event happened. Offering this explanation spares the explainee from this work, a form of computational kindness (Christian & Griffiths, 2016). Exploring this hypothesis is a fruitful direction for future research.

References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Christian, B., & Griffiths, T. (2016). *Algorithms to live by: The computer science of human decisions*. Macmillan.

Davis, Z. J., & Rehder, B. (2020). A process model of causal reasoning. *Cognitive Science*, *44*(5), e12839.

De Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, *47*, 1–12.

Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological Review*.

Gerstenberg, T., & Icard, T. (2020). Expectations affect physical causation judgments. *Journal of Experimental Psychology: General*, *149*(3), 599.

Hagmayer, Y., Sloman, S. A., Lagnado, D. A., & Waldmann, M. R. (2007). Causal reasoning through intervention. In A. Gopnik & L. Schulz (Eds.), *Causal learning: Psychology, philosophy, and computation* (pp. 86–100, chapter = 6). Oxford University Press.

Halpern, J. Y. (2016). *Actual causality*. MiT Press.

Henne, P., Niemi, L., Pinillos, Á., De Brigard, F., & Knobe, J. (2019). A counterfactual explanation for the action effect in causal judgment. *Cognition*, *190*, 157–164.

Hesslow, G. (1988). The problem of causal selection. In D. J. Hilton (Ed.), *Contemporary science and natural explanation: Commonsense conceptions of causality.* New York University Press. doi: 10.1086/355318

Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition*, *161*, 80–93.

Konuk, C., Goodale, M. E., Quillien, T., & Mascarenhas, S. (2023). Plural causes in causal judgment. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *45*(45).

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. doi: 10.18637/jss.v082.i13

Lagnado, D. A. (2021). *Explaining the evidence: How the mind investigates the world*. Cambridge University Press.

Lagnado, D. A., Gerstenberg, T., & Zultan, R. (2013). Causal responsibility and counterfactuals. *Cognitive Science*, *37*(6), 1036–1073.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in Cognitive Sciences*, *10*(10), 464–470.

Lucas, C. G., & Kemp, C. (2015). An improved probabilistic account of counterfactual reasoning. *Psychological Review*, *122*(4), 700.

Meder, B., & Mayrhofer, R. (2017). Diagnostic causal reasoning with verbal information. *Cognitive Psychology*, *96*, 54–84.

Meteyard, L., & Davies, R. A. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, *112*, 104092.

Morris, A., Phillips, J., Gerstenberg, T., & Cushman, F. (2019). Quantitative causal selection patterns in token causation. *PloS one*, *14*(8), e0219704.

Pearl, J. (2009). *Causality*. Cambridge university press.

Quillien, T. (2020). When do we think that x caused y? *Cognition*, *205*, 104410.

Quillien, T., & Barlev, M. (2022). Causal judgment in the wild: evidence from the 2020 us presidential election. *Cognitive Science*, *46*(2), e13101.

Quillien, T., & Lucas, C. G. (2023). Counterfactuals and the logic of causal selection. *Psychological Review*.

Sloman, S., & Lagnado, D. A. (2004). Causal invariance in reasoning and learning. *Psychology of Learning and Motivation*, *44*, 287–326.

Woodward, J. (2021). *Causation with a human face: Normative theory and descriptive psychology*. Oxford University Press.

Zultan, R., Gerstenberg, T., & Lagnado, D. A. (2012). Finding fault: Causality and counterfactuals in group attributions. *Cognition*, *125*(3), 429–440.