Plural causes in causal judgment

Can Konuk (konukcan@gmail.com)
Institut Jean Nicod, Département d'études cognitives,
Ecole Normale Supérieure, EHESS, CNRS, PSL University
29 rue d'Ulm, 75005 Paris, France

Michael Goodale (michael.goodale@mail.mcgill.ca)
Institut Jean Nicod, Département d'études cognitives,
Ecole Normale Supérieure, EHESS, CNRS, PSL University
29 rue d'Ulm, 75005 Paris, France

Tadeg Quillien (tadeg.quillien@gmail.com)
School of Informatics, University of Edinburgh
10 Crichton Street, Edinburgh EH8 9AB, Scotland, United Kingdom

Salvador Mascarenhas (salvador.mascarenhas@ens.fr)
Institut Jean Nicod, Département d'études cognitives,
Ecole Normale Supérieure, EHESS, CNRS, PSL University
29 rue d'Ulm, 75005 Paris, France

Abstract

Causal selection is the process whereby people decide which of several events responsible for a realized outcome should be considered as "the cause" of that outcome. A theory of causal selection requires a definition of the relevant candidates to be considered for selection. So far, the psychological literature has operated on the implicit premise that the only relevant candidates for causal selection are individual variables, corresponding to the distinct nodes of a causal network. Instead, we argue that causal judgment can recognize *plural causes*, featuring more than one variable. We provide evidence for the psychological relevance of plural causes by showing: (a) that plural cause judgments are influenced by the same factors that have been proven to influence causal selection judgments in general; (b) that this influence cannot be explained away by assuming that participants estimate the strength of plural cause simply by combining the strength of its individual constituents.

Keywords: actual causation; plurals ; causal selection ; counterfactual theories of causation

Introduction

Causal selection is the process by which people are driven to say that an outcome happened 'because of' a given event, or that the event is 'the cause' of that outcome. When a forest catches fire after a lightning strike, for example, people tend to say that the lightning bolt was the cause of the fire, and not mention the presence of oxygen in the air, although the latter was no less indispensable for the fire to occur.

The outcome of causal selection judgments depends on the initial pool of candidates that are considered to begin with. For the lightning bolt to be viewed as causing the fire, the variable 'ligthning' must first be flagged by the mind as a relevant candidate for causal selection. We argue here that the extant psychological literature on causal selection has had a blind spot regarding that initial pool of candidates: it operates on the implicit premise that the only relevant variables for causal selection are individual variables, corresponding to distinct nodes in the relevant network of causes.

Instead, we argue that causal selection judgments can recognize *plural causes*, featuring more than one variable, as when we say that 'the dryness of the season and the strength of the wind' caused the uncontrollable spread of the fire.

That causal cognition admits causes featuring several variables is not itself a new idea. In causal inference, researchers have studied how people infer conjunctive causes—factors that act in concert to produce an effect (Novick & Cheng, 2004). The notion of a plural cause also plays a role in some theories of actual causation, such as Halpern's (2015). To our knowledge, however, plural causes have not been explored for the case of causal selection.

Here we empirically establish the psychological reality and non-triviality of plural causes. We show that people's judgments about plural causes are sensitive to the prior probability of events—a key signature of human causal selection judgment. More importantly, we show that judgments about the causal strength of a combination of causes (e.g. '*A* & *B*') cannot simply be reconstructed from judgments about the causal strength of the individual variables ('*A*' and '*B*'), providing evidence that people see plural causes as full-fledged candidates for causal selection.

Moreover, our findings have implications for studying the role of counterfactual reasoning in causal judgment. First, we show that models based on counterfactuality can be straightforwardly extended to make non-trivial predictions consistent with our findings. Second, while our goal here is not compare alternative counterfactual models, the new empirical domain of plural causal selection has great potential for generating contrasting predictions, a topic we are exploring in ongoing work.

Theoretical background

A large body of evidence supports the idea that much of human thought is based on causal models of the world (Chater

In M. Goldwater, F. K. Anggoro, B. K. Hayes, & D. C. Ong (Eds.), *Proceedings of the 45th Annual Conference of the Cognitive Science Society*.

& Oaksford, 2013; Sloman & Lagnado, 2015; Gerstenberg & Tenenbaum, 2017; Pearl & Mackenzie, 2018). In causal cognition research, these models are represented using formalisms such as Causal Bayes Nets or Functional Causal Models: these formal systems represent aspects of the world with variables, causal relationships between these variables, and probability distributions (Pearl, 2000).

Suppose for example that I get a stomach ache shortly after having eaten a chocolate cake, a lemon pie, and a panna cotta from a plate of desserts. A causal model of this situation would feature one variable for each of the causes of my stomach ache (i.e. one variable each for 'eating the chocolate cake', 'eating the lemon pie', and 'eating the panna cotta') as well as a variable for the effect ('having a stomach ache'). The model also specifies a functional relationship between the variables, for example representing the fact that one develops stomach issues after eating too much.

As a consequence of this representational format, it seems natural to think of the candidates for causal selection as particular realizations of the individual variables. Indeed, a striking feature of the psychological literature is that causal selection judgments are only ever queried at the level of *singular* variables—the variables that stand as clearly distinct nodes in the causal system of interest. When participants are presented with a situation where, for example, the outcome depends on the conjunction of three different events A, B, or C, they are never asked to what extent the *plural event* 'A & B' was the cause of the outcome. They are only asked about each individual event separately.

Intuitively, however, causal explanations that mention combinations of variables can also be appealing. In our example above, saying that I got a stomach ache 'because I ate the entire plate of desserts' might appear to be a better explanation than 'because I ate the chocolate cake'. Note that allowing for many variables to feature in causal explanations does not eliminate the need for causal *selection*: for example one might feel that 'because I ate the entire plate of desserts' is still the best causal explanation for my stomach ache even if it leaves out the one blueberry that I also ate afterward.

Despite the intuitive naturalness of plural causes, current psychological theories of causal selection are silent about them. In the next paragraphs we distinguish two possible hypotheses about how the mind computes the causal strength of a plural cause, and outline our strategy for arbitrating between these hypotheses.

Linearity vs. holistic computation

People might simply make judgments about a plural cause by adding up, or averaging, the causal strength estimates for its constituent variables. For example, to compute how much they agree that 'eating the chocolate cake, the lemon pie and the panna cotta caused the stomach ache', people first compute their agreement with 'eating the chocolate cake caused the stomach ache', 'eating the lemon pie caused the stomach ache', etc, and then they combine the causal strength of each individual variable. We will call this the *linear combination* hypothesis. This hypothesis is deflationary with respect to the psychological reality of plural causes: people can make plural cause judgments, but they cobble them together from more primitive representations of causal strength at the level of individual variables.

In contrast, we consider the possibility that plural causal judgments are the output of a holistic computation. Under that account, causal judgments about a combination of variables might significantly diverge from the causal judgments for its constituent individual variables. We are not committed to one particular specification of this computation, but below we will consider one operationalization, in the context of a counterfactual account of causal judgment (Quillien & Lucas, 2023).

The Counterfactual Effect Size Model

A large body of research suggests that human causal judgment relies on counterfactual reasoning (Morris et al., 2018; Halpern & Hitchcock, 2015; Spellman, 1997; Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2021; Icard, Kominsky, & Knobe, 2017; Quillien, 2020; Quillien & Lucas, 2023). Recent counterfactual theories have been largely successful at predicting people's causal judgments across a wide range of tasks (Quillien & Barlev, 2022; Henne, Kulesza, Perez, & Houcek, 2021; Kirfel & Lagnado, 2021; Morris, Phillips, Gerstenberg, & Cushman, 2019; O'Neill, Henne, Bello, Pearson, & De Brigard, 2022; Henne, Niemi, Pinillos, De Brigard, & Knobe, 2019; Gerstenberg & Icard, 2020). Here we focus on a recent counterfactual account that parsimoniously captures many empirical features of human causal judgment. According to Quillien and Lucas's (2023) Counterfactual Effect Size Model (CESM), causal selection is the result of the following process:

1. First, sample a large number of counterfactuals to the present situation. Across counterfactual worlds, each of the relevant causal variables is sampled with a frequency that depends on: its prior probability (a tractable proxy for normality), and the value that the variable takes in the real world. That is, an event is more likely to be represented in counterfactual worlds if it did happen in the real-world, and was likely to happen to begin with.

2. Second, compute the causal impact of that variable for the relevant outcome by looking at their correlation across counterfactuals thus sampled. The measure of that impact attaches a 'causal strength' score to each of the variables. From there, causal selection judgments are understood as a direct result of the ranking implicit in that scoring process.

We can consider two ways that the CESM might handle plural causes, each corresponding to one of the hypotheses we laid out above. Under the linear combination hypothesis, people compute the causal strength of the plural 'A & B' for outcome E by first computing the causal strength of A and the causal strength of B. This means to compute the cross-counterfactual correlations between A and E and between B

and E. Then they combine these scores in some linear fashion, for example by averaging them. Under the holistic computation account, people directly compute the correlation between 'A & B' and E across counterfactuals. That is, they track the value of a composite variable that is true in worlds where both A and B are true, and false otherwise. Then they compute the correlation between this variable and the outcome. The plural cause has a high causal strength to the extent that this correlation is high.

The present study

Here we present (to our knowledge) the first empirical investigation of plural causes in the context of causal selection. Our study has the following goals. First, if plural causes are processed as genuine causes by the mind, factors that are known to affect causal-selection judgment should influence judgment about a plural cause. Notably, the *probability* of an event is known to affect judgments about whether that event caused an outcome (Morris et al., 2019). We expect analogous patterns of effect for plurals: varying the probability of events should affect causal judgments about whether a conjunction of these events caused the outcome.

Second, we aim to rule out the linear combination account. Evidence of *non-linearity* in people's causal judgments would constitute stronger evidence for the psychological reality of plural causes in human causal selection. We design a situation where the holistic version of the CESM predicts that the causal strength of plural variables will *not* be a linear combination of the score of individual variables. We compare its predictions to those of a null model that tries to predict the score of plural causes through a linear combination of the score of individual variables.

Experiment

Design and materials

We adapted a paradigm developed by Quillien and Lucas (2023). Participants made judgments about a game of chance, in which one randomly draws balls from a set of urns, and wins by getting enough colored balls (see Figures 1 and 2 for illustrations). Participants observed a fictitious player draw a colored ball from each of three urns (labelled A, B and C) and win the game as a result. Then they were asked to make a causal judgment about each singular cause (e.g. whether getting a colored ball from urn A caused the player to win the game), and about each pair (e.g. whether getting a colored ball from urns A and B caused the player to win the game). For exploratory purposes, we also asked participants to make a causal judgment about the triple (getting a colored ball from A, B and C).

We manipulated the prior probability of each outcome (within participants) by varying the proportion of colored balls in each urn (yielding probabilities of 0.05, 0.5, and 0.95; see Figure 1). We will refer to the three different urns as the low-, intermediate-, and high-probability urns, respectively. The rules of the game were that the player wins if they get two colored balls or more.

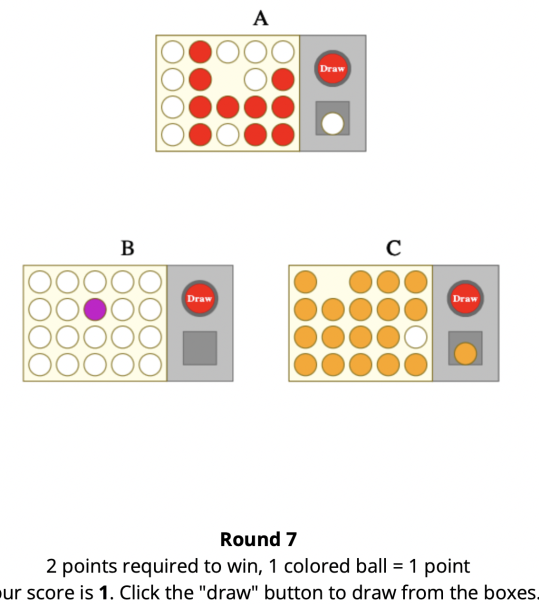

Figure 1: A round of the game where participants are invited to draw from all three urns. Here the participant drew a white ball from Urn A, a colored one from Urn C, and has yet to draw from Urn B.

Predictions

This paradigm provides a context where the linear and the holistic extensions of the Counterfactual Effect Size Model we outlined above make clearly different predictions.

The CESM predicts that participants' singular causal strength estimates should follow a particular ranking: intermediate probability urn > low probability urn > high probability urn (Figure 3). This is because, across possible counterfactual alternatives to what happened, there is a high correlation between getting a colored ball from the intermediate-probability urn and winning the game.

In the linear extension to CESM, where participants simply combine the causal strength of individual variables to make plural cause judgments, they should consider that the pair 'low and intermediate' should have greater-than-or-equal causal strength to the pair 'high and intermediate', because the singular 'low' has higher causal strength than 'high'.

In contrast, if participants judge the causal strength of plurals via a holistic CESM computation, they should rate the pair 'intermediate and high' as highest. For, across possible counterfactuals, there is a high correlation between getting a colored ball from these two urns and winning the game. Intuitively, since drawing a colored ball from the low-probability urn is rare, most worlds where the player wins the game will be worlds in which they do so by getting a colored ball from the other two urns.

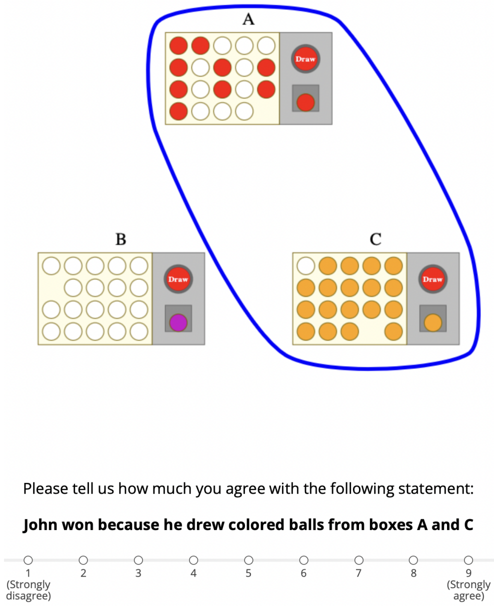

Figure 2: The prompts presented to participants next to the fictitious player's draws.

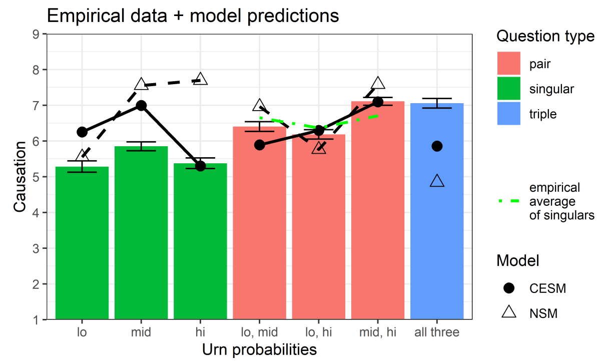

Figure 3: Mean responses by question and predictions for each model. The error bars represent the standard error of the mean. As one can see here (dotted green line), the 'linear combination' theory predicts that the score of the 'lo, mid' and 'mid, hi' pairs should be equivalent, when in fact we see a significant difference between them, in accordance with the 'holistic' versions of two counterfactual models: the Counterfactual Effect Size Model (Quillien & Lucas, 2023) and the Necessity and Sufficiency Model (Icard et al., 2017).

Procedure

Participants first completed ten rounds of the game themselves, presented with urns as in Figure 1. The draws were pseudo-randomized in order to get participants to internalize the probabilities associated with each urn and their link with the outcome.

Then participants were shown the outcome of a round of the game played by another fictitious player, who drew a colored ball from *all three* urns (thereby winning with 3 points). They were asked to rate the causal strength of each individual draw, as well as that of every combination of (two or three) draws for the winning outcome, on a Likert scale from 1 to 9 (strongly disagree to strongly agree), as in Figure 2. For the singulars, participants were asked to rate their agreement with the statement 'John won because he drew a colored ball from box [urn]'; for the plurals they rated their agreement with 'John won because he drew colored balls from boxes [urn_1] and [urn_2]'.

Each question was displayed on a separate page, next to the urns that displayed the outcome of the fictitious player's draw (see Figure 2). The letters indexing the urns, as well as the colors of the balls, were randomized across participants to avoid confounding (but were kept the same within a participant). Half of the participants were asked about the singulars first, and then about the pairs. The other half were asked about the pairs, and then about the singulars. All participants were asked about the triple at the end. Within one class of questions (e.g. questions about singulars) the order of presentation of urns was randomized.

Finally, participants completed a brief demographic questionnaire, were thanked for their participation and re-directed to Prolific for payment. The experiment was coded in the `jsPsych` library (De Leeuw, 2015), with custom plugins for displaying urns developed in our lab. All data and `R` code for analysis are available on the Open Science Framework at `https://osf.io/3th7a/`.

Participants

We recruited 400 participants from all English-speaking countries from Prolific. We excluded from subsequent analysis 44 participants that failed to answer either one of two elementary comprehensions questions that checked their understanding of the rules of the game, leaving a total of 356 participants.

Results

We first report analyses using standard statistical tests. Then we report the fit of computational models of causal judgment.

Model-free results

Prior probability affects both singular and plural causal judgments. Results are plotted in Figure 3. We ran two two-factor repeated-measure ANOVAs, one for each main type of cause queried ('singulars' and 'pairs'), using urn probabilities and order of presentation as predictor variables, and participants' responses as the dependent variable. Results are presented in Tables 1 and 2. There was a main effect of prior probability on participants' causal judgments, for singular as well as for plural causes ($p < 0.02$ in both cases), consistent with our expectation that participants' judgments for plural causes should be sensitive to probabilities just like for other actual cause judgments.

The order of presentation (whether singulars were presented before or after plurals) also had a significant effect

Factors	Mean Sq	*F* value	*p*-value
Probabilities of the urns	32.98	4.492	< 0.012
Order of presentation	138.81	18.904	1.51e-05
Probabilities:order	3.18	0.433	0.6488

Table 1: ANOVA for singular urns.

Factors	Mean Sq	F value	*p*-value
Probabilities of the urns	83.02	14.646	5.32e-07
Order of presentation	21.75	3.837	0.0504
Probabilities:order	2.30	0.406	0.6662

Table 2: ANOVA for pairs of urns.

Factors	Mean Sq	F value	*p*-value
lo	100.69	17.743	< 0.00001
mid	20.80	3.664	0.05586
hi	15.93	2.808	0.094
lo:mid	6.64	1.169	0.2798
lo:hi	0.41	0.073	0.7872
mid:hi	46.64	8.219	0.00423

Table 3: Results of the ANOVA: ESTIMATE FOR PAIRS $\sim$ EST. FOR SINGULAR 1 $\times$ EST. FOR SINGULAR 2.

Models	LogLik	Df	χ^2	*p*-value	BIC
Means sing	-891.89	3			1804.709
+ Question	-878.13	5	27.53	1.052e-06	1791.126

Table 4: Comparison between the two models, the linear combination model of plurals (MEANS OF SINGULARS + INTERCEPT), and the MEANS OF SINGULARS + QUESTION model.

($p < 0.001$) on the ratings for singulars: singular causal judgments were lower when presented after the plurals. There was however no interaction effect between urns probability and order of presentation, suggesting that the impact of probability on causal estimates did not vary depending on the order in which questions were asked. Therefore we drop this variable (order of presentation) from later analyses.

The causal strength of plural causes is not a linear combination of the causal strength of individual variables. The pattern of responses for singular variables replicated the patterns obtained by Quillien and Lucas (2023). Judgments for the intermediate-probability urn were higher than judgments for the low-probability urn, $t(315.41) = -2.70$, $p = 0.007$, and the high-probability urn, $t(325.85) = -2.08$, $p = 0.038$. The difference between the low and high probability urns was not significant, $t(350.59) = -0.63$, $p = 0.53$.

We can use these results to test the 'linear combination' hypothesis, according to which participants derive their plural cause strength estimates by adding up or averaging their estimates for the individual variables that compose a given plural. If this were correct, participants should give the same causal strength estimate for the two plural causes 'low urn and intermediate urn' and 'intermediate urn and high urn', since their estimates for the singular causes 'low probability urn' and 'high probability urn' are not significantly different from each other. By contrast, the holistic CESM predicts a sharp difference between these two kinds of plurals, with the plural cause 'intermediate and high urn' being rated higher (Figure 3).

Consistent with the CESM, judgments about the 'intermediate, high' pair were higher than for the 'lo, intermediate' pair, $t(355) = -4.67$, $p < 0.001$, and higher than for the 'lo, high' pair, $t(355) = 6.858$, $p < 0.001$. In slight deviation from the CESM predictions however, judgments for the 'low, intermediate' pair were higher than for the 'lo, high' pair, $t(355) = 2.3691$, $p = 0.02$; see Figure 3.

We conducted two other analyses to further rule out the linear combination model. First, we ran a one-way repeated-measure ANOVA, predicting judgments about the pairs ('low and intermediate', 'intermediate and high', and 'low and high') from judgments about the singulars ('low', 'intermediate', and 'high'), as well as their interactions, as within-participant factors. Each plural pair was regressed only on the values of the two singulars that it comprised.

The linear combination theory predicts that there should be no significant interaction: a participant's causal judgment for a given *singular* variable should have the same impact on every plural cause in which it features. One's estimate for the singular 'intermediate urn', for example, should have an equal impact on one's estimate for 'intermediate and high' and for 'low and intermediate'.

We find evidence against the 'linear combination' theory (Table 3). There was a significant interaction between the 'intermediate' and 'high' probability urns, $p = 0.004$. In addition, the main effects of the singular judgments were not significant, for all but the 'low' probability urn.

Second, we fitted linear multilevel regression models on participants responses for pairs. Specifically, we compared the predictive performance of two different models on participants' plural cause estimates. The first one used as predictor the average of the two singular cause estimates for the variables contained in a given plural (computed on a per-participant basis), plus a random intercept. The second model also included the question asked (that is, the specific plural being queried) as predictor. A likelihood ratio test shows that adding question as a predictor significantly improves the fit of the model, $\chi^2(5) = 27.53$, $p < 0.001$ (Table 4). Again, this is inconsistent with the linear combination account.

Computational modeling

We computed the predictions of two recent counterfactual models of causal selection, the CESM (presented in the Introduction) and the Necessity and Sufficiency Model (Icard et al., 2017). For space considerations we cannot fully describe the formal models here, but our implementation follows the one given in Quillien and Lucas (2023). In particular we use the same value they used for the 'stability' parameter s ($s = 0.73$ and $s = 0.15$ for the CESM and NSM respectively).

We constructed two versions of both the NSM and CESM for plural causes. The 'additive' version of a model first computes the causal strength of each individual cause in a conjunct, and then averages them. For example, the additive version of the CESM computes the causal strength of 'low, high' by computing the (cross-counterfactual) correlation between Low and the outcome, computing the correlation between High and the outcome, and then averaging these two correlations together. The holistic versions of the models (that we will simply call "NSM" and "CESM") assume that people compute the causal strength of the conjunct directly. For instance, the holistic version of the CESM computes the causal strength of 'low, high' by computing the correlation between a binary variable LH (which is 1 if both Low and High are 1, and 0 otherwise) and the outcome.

Results Model predictions are plotted in Figure 3. We tested the fit of each model to the data, excluding the judgments for the triples (which were exploratory). For each model we computed a multilevel regression with model judgment as predictor, participant-specific random intercepts, and human causal judgment as dependent variable, and extracted the BIC of the regression. We find that the best-fitting model is a holistic model (the holistic version of the CESM, BIC = 9945), providing further evidence against the linear combination theory. All other counterfactual models had a better fit than a random baseline (random baseline, BIC = 9987; additive CESM, BIC = 9985; NSM, BIC = 9985; additive NSM, BIC = 9978).

Discussion

We find evidence that when people make a judgment about whether A and B caused an outcome, their judgments track the correlation between the conjunction of A and B and the outcome, across counterfactuals. Concretely, in our experiment, winning the game is in general strongly associated with getting a ball from both the intermediate- and high-probability urns, and people judged that combination of events to be highly causal. Importantly, this effect is inconsistent with a simpler account, according to which people's judgments about a plural are cobbled together from their causal intuitions about each individual variable in the plural.

As such, our results are in general consistent with the predictions of recent counterfactual models of causal selection (Icard et al., 2017; Quillien, 2020; Quillien & Lucas, 2023), augmented with the assumption that people judge plural causes in a holistic manner. In particular, the Counterfactual Effect Size Model (Quillien, 2020; Quillien & Lucas, 2023) provides a good account of the data.

There were however two interesting effects that the model does not capture. First, participants found the plurals more appealing than the singulars. Participants might have felt that plurals provide more exhaustive *descriptions* of the event: they give more complete information about what happened, in addition to why it happened. This finding suggests an interesting tension between two potential desiderata of causal judgment: highlighting the variable(s) that were most causally important to the outcome, and providing an exhaustive list of the causal factors. We also find that this effect is accentuated when singulars are presented after plurals. Making judgments about plurals first might highlight to participants the descriptive incompleteness of the singulars.

Second, participants had a slight tendency to prefer the pair 'low, intermediate' to the pair 'low, high', while the CESM tends to predict the inverse tendency. One possible explanation is that, despite the limitations of the linear combination account outlined above, the causal strength of individual variables might still influence causal judgments about plurals to some extent. The 'intermediate' singular was judged as more causal than the 'high' singular, and some of that effect might have carried over to the corresponding pairs.

General discussion and conclusion

The current study is a preliminary demonstration that, to the mind, causes are more than the sum of their parts. Judgments about plural causes are affected by the prior probability of their constituent variables, but cannot be derived from the causal strength of these individual variables. Our results are consistent with simple extensions of extant counterfactual models of causal selection. At the same time, our findings raise new issues about the psychology of causation, which point to opportunities for extending existing theories.

We conclude by considering a normative question: *should* plural causes be processed in the holistic manner we have documented? We argue that the current model has one important advantage: it produces judgments that are invariant with respect to the granularity of the causal model. The question of which variables count as singular, and which ones count as plural, is only relative to the level of granularity with which we choose to look at the situation (Beckers & Halpern, 2019; Kinney & Lombrozo, 2022). A single variable like 'eating dessert' can for example be broken down into multiple subevents like 'having a chunk of cake', 'having another chunk of cake', 'having a cherry', etc. Intuitively, causal attribution should not be dramatically affected by the granularity with which we construe the situation. Our model respects this intuitive criterion: it makes the same judgment about the causal contribution of eating the cake, whether we choose to conceive of it as one event or as a conjunction of events.

Acknowledgments

We thank Thomas Icard, Nicolas Navarre, and the audiences of the LANG-REASON seminar at Ecole Normale Supérieure for valuable comments. This work was supported by Agence Nationale de la Recherche grants ANR-18-CE28-0008 (LANG-REASON; PI: Mascarenhas) and EURIPANR-17-EURE-0012 (Investments d'Avenir Program), and by

Ecole Doctorale Frontières de l'Innovation en Recherche et Education—Programme Bettencourt.

References

Beckers, S., & Halpern, J. Y. (2019). Abstracting causal models. In *Proceedings of the aaai conference on artificial intelligence* (Vol. 33, pp. 2678–2685).

Chater, N., & Oaksford, M. (2013). Programs as causal models: Speculations on mental programs and mental representation. *Cognitive Science*, *37*(6), 1171—1191. doi: 10.1111/cogs.12062

De Leeuw, J. R. (2015). jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior research methods*, *47*(1), 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*, *128*(5), 936–975. doi: 10.1037/rev0000281

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

Gerstenberg, T., & Tenenbaum, J. B. (2017). *Intuitive theories*. Oxford University Press.

Halpern, J. Y. (2015). A modification of the Halpern-Pearl definition of causality. *arXiv preprint*. doi: 10.48550/arXiv.1505.00162

Halpern, J. Y., & Hitchcock, C. (2015). Graded causation and defaults. *The British Journal for the Philosophy of Science*, *66*(2). doi: 10.1093/bjps/axt050

Henne, P., Kulesza, A., Perez, K., & Houcek, A. (2021). Counterfactual thinking and recency effects in causal judgment. *Cognition*, *212*. doi: 10.1016/j.cognition.2021.104708

Henne, P., Niemi, L., Pinillos, A., De Brigard, F., & Knobe, J. (2019). A counterfactual explanation for the action effect in causal judgment. *Cognition*, *190*, 157–164. doi: 10.1016/j.cognition.2019.05.006

Icard, T. F., Kominsky, J. F., & Knobe, J. (2017). Normality and actual causal strength. *Cognition*, *161*, 80–93. doi: 10.1016/j.cognition.2017.01.010

Kinney, D., & Lombrozo, T. (2022). Evaluations of causal claims reflect a trade-off between informativeness and compression. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 44).

Kirfel, L., & Lagnado, D. (2021). *Changing minds — epistemic interventions in causal reasoning.* (on PsyArXiv) doi: 10.31234/osf.io/db6ms

Morris, A., Phillips, J., Gerstenberg, T., & Cushman, F. (2019). Quantitative causal selection patterns in token causation. *PLoS ONE*, *14*(8). doi: 10.1371/journal.pone.0219704

Morris, A., Scott-Philips, J., Icard, T. F., Knobe, J., Gerstenberg, T., & Cushman, F. (2018). *Judgments of actual causation approximate the effectiveness of interventions.* (Psy ArXiv) doi: 10.31234/osf.io/nq53z.

Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal influence. *Psychological Review*, *111*(2), 455–485. doi: 10.1037/0033-295X.111.2.455

O'Neill, K., Henne, P., Bello, P., Pearson, J., & De Brigard, F. (2022). Confidence and gradation in causal judgment. *Cognition*, *223*. doi: 10.1016/j.cognition.2022.105036

Pearl, J. (2000). *Causality: models, reasoning, and inference*. USA: Cambridge University Press. doi: book/10.5555/331969

Pearl, J., & Mackenzie, D. (2018). *The book of why: The new science of cause and effect*. Basic Books, Inc. doi: 10.5555/3238230

Quillien, T. (2020). When do we think that x caused y? *Cognition.*, *205*. doi: 10.1016/j.cognition.2020.104410

Quillien, T., & Barlev, M. (2022). Causal judgment in the wild: Evidence from the 2020 u.s. presidential election. *Cognitive Science*, *56*(2). doi: 10.1111/cogs.13101

Quillien, T., & Lucas, C. G. (2023). Counterfactuals and the logic of causal selection. Forthcoming in *Psychological Review*. doi: 10.31234/osf.io/ts76y

Sloman, S. A., & Lagnado, D. A. (2015). Causality in thought. *Annual review of psychology*, *66*, 223–247.

Spellman, B. A. (1997). Crediting causality. *Journal of Experimental Psychology: General*, *124*(4), 323—348. doi: 10.1037/0096-3445.126.4.323