UC Merced
Proceedings of the Annual Meeting of the Cognitive Science Society

Title
Causation by Ignorance

Permalink
https://escholarship.org/uc/item/79t8w08k

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 43(43)

ISSN
1069-7977

Authors
Kirfel, Lara
Lagnado, David

Publication Date
2021

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Causation by Ignorance

Lara Kirfel (ucjulki@ucl.ac.uk)

David Lagnado (d.lagnado@ucl.ac.uk)
Department of Experimental Psychology, University College London, 26 Bedford Way
London, WC1H 0AP England

Abstract

Epistemic states, what an agent knows or beliefs, play a crucial role in people's moral evaluations of the agent's actions. Whether and to what extent epistemic states also influence an agent's perceived causal contribution to an outcome remains the subject of debate. In three experiments, we investigate people's causal and counterfactual judgments about ignorant causal agents. We find that agent's epistemic states, the conditions of their ignorance as well as their epistemic actions influence how causal an agent is perceived, but also the kind of counterfactual alternatives people consider. We take these findings to indicate the crucial role of epistemic states in causal cognition and counterfactual models of causation.

Keywords: causal judgment; counterfactual reasoning; epistemic states; ignorance; blame

Introduction

If Dr. Jones unknowingly prescribes her patient a drug that causes unforeseen side effects, we will likely attenuate our blame response in light of her ignorance about the consequences of her action. But how do we judge her causal role in this scenario? Recent studies in causal cognition find evidence that epistemic states like knowledge or ignorance influence how causal people perceive an agent for the outcome of their action (Lagnado & Channon, 2008; Lombrozo, 2010; Hilton, McClure, & Moir, 2016). Agents lacking knowledge (Gilbert, Tenney, Holland, & Spellman, 2015) or foreseeability of the consequences of their actions (Lagnado & Channon, 2008) are judged to be less of a cause for these outcomes. In causal chains, the causality of knowing agents is rated higher than that of ignorant ones (Hilton et al., 2016; McClure, Hilton, & Sutton, 2007; Lombrozo, 2010). If the proximal cause is a human action, the agent is judged as more causal if the agent was aware of the causal opportunity created by prior events (Hilton et al., 2016). Likewise, people's preference for abnormal actions as causes has been shown to be mediated by the agents' knowledge states (Kirfel & Lagnado, 2021; Samland & Waldmann, 2016).

Why do epistemic states matter for causal assessments? Given the essential role of mental states for moral judgments, epistemic influences on causal attributions have been argued to be influenced or biased by moral evaluations (Alicke, Rose, & Bloom, 2012; Alicke & Rose, 2012). On the other hand, it has been suggested that the influence of epistemic states might uncover something more fundamental about how people judge about the causality of agents (Lombrozo, 2010). Gilbert et al. (2015) show that the difference in people's causal attributions to knowing vs. ignorant agents is mediated by counterfactual thinking. In their studies, they find that in case of agents who know about the (negative) consequences of their actions, participants generate more counterfactuals about ways the outcome could have been different that the actor could control. Notably, their results suggest that the influence of epistemic states is driven by whether, and especially *what* counterfactuals people consider (Spellman & Gilbert, 2014).

Counterfactual Causation and Epistemic States

According to counterfactual theories of causation, C is a cause of E if E is counterfactually dependent on C, that is, E would not have happened in the absence of C (Lewis, 2013; Woodward, 2007). Counterfactual dependence is assessed in terms of hypothetical interventions (Woodward, 2007; Halpern, 2016; Pearl, 2009) or mental simulations (Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2020) over causal candidate variables, often represented in form of a *do*-operator that sets a certain variable to a certain value $do(X = x)$ (Pearl, 2009). Halpern (2016) extends to this framework to the evaluation of counterfactual dependence under different "contingencies", i.e. non-actual possible worlds in which certain background variables are different.

Counterfactual models of causation have been shown to accurately capture the structural aspects that influence people's causal judgments about a cause (Gerstenberg et al., 2020). However, these theories seem to fail to account for a fundamental aspect in people's causal thinking: the difference in people's causal judgments about knowing vs. ignorant causal agents. In the case of social or agent causation, it is often assumed that counterfactual interventions target an agent's action (Halpern, 2016; Gerstenberg et al., 2020), that is, testing whether the undoing of the Doctor giving the drug leads to a difference in the outcome, the patient's health problems. Such a counterfactual dependence test however is insensitive towards the agent's epistemic states, as it would render the doctor a cause of the outcome, irrespective of whether the doctor knew about the side effects of the drug, or not.

Hypotheses Drawing on previous research by Gilbert et al. (2015), we aim to explore the question if the influence of epistemic states on causal judgments can be accounted for by counterfactual theories of causal reasoning. Here, we put forward the hypothesis that the intervention that people perform in counterfactual reasoning about ignorant agents targets the agent's epistemic state and epistemic conditions. Rather than assessing whether Dr Jones' not prescribing the drug would have made a difference to the outcome, intuitively we might want to change her epistemic state from ignorance to knowledge first, or even think about potential ways in which she could have acquired the relevant knowledge. In three experiments, we aim to test whether people's causal judgments are sensitive not only to the agent's epistemic state, but also to the agent's "epistemic actions", i.e. their knowledge-seeking behaviour. Crucially, we hypothesise that these epistemic conditions are reflected in people's counterfactual thinking.

1. **Hypothesis 1**
   (a) *Causal Judgment:* Ignorant agents are judged as less causal than knowing agents.
   (b) *Counterfactual Reasoning:* If the causal agent is ignorant, people intervene on epistemic states, rather than on causal actions.
2. **Hypothesis 2**
   (a) *Causal Judgment* Ignorant agents who could have changed their epistemic states are judged more causal than those who could not.
   (b) *Counterfactual Reasoning* If the causal agent is ignorant, people intervene on the agent's *epistemic actions*, rather than on causal actions.

Hypothesis 1 aims to test more generally whether people intervene on an epistemic states. Comparing a knowledgeable vs. an ignorant causal agent, we predict that the latter is judged less causal, and that counterfactual reasoning will target the agent's epistemic state (*Experiment 1: Knowledgeable vs. Ignorant Agents*). Hypotheses 2 makes predictions for cases in which a causal agent is ignorant, but could — by their own epistemic actions — have changed their state of ignorance and acquired knowledge. We will test Hypothesis 2 for two different cases of epistemic action conditions. In the most basic case, an agent is ignorant and either could or could not have acquired knowledge by their own actions (*Experiment 2: Externally vs. Self-Caused Ignorance*). In such a case, we predict that the agent who could have changed their epistemic state will be judged as more causal, and that people will intervene on this' agent's epistemic (non-)action. Finally, we turn back to the idea that counterfactual dependence is assessed under different contingencies (*Experiment 3: Epistemic Actions under Different Contingencies*). We predict that an agent whose epistemic action did not lead them to acquire knowledge in the actual world, but would have led them to acquire knowledge under different circumstances, is judged less causal than one who would remain ignorant in both actual and possible world.

Experiment 1

In Experiment 1, we aimed to investigate people's causal judgment and counterfactual reasoning about an agent who knows vs. does not know the consequences of their action.

Participants and Design

We recruited 145 participants on Amazon Mechanical Turk. 23 participants were excluded for failing one or more of the four comprehension check questions, and one participant was excluded for providing a non-sensical counterfactual response (see below, leaving a final sample size of $N = 121$ (M_{age} = 38.42, SD_{age} = 11.15, N_{female} = 40). We adopted a 2 knowledge (knowledge vs. no knowledge) × 3 scenario ("hospital" vs. "garden" vs. "bakery") design. The 'Knowledge' condition was manipulated as within-participants contrast in order to allow for the contrastive nature of counterfactual reasoning (Schaffer, 2005; McGill & Klein, 1993). 'Scenario' was manipulated between participants.

Materials and Procedure

Participants read both the 'knowledge' as well as the 'no knowledge' condition of one of the three scenarios ("hospital", "garden", "bakery") in randomised order. All three scenarios follow the same content structure: As part of their work, an agent usually applies a certain a product ("medical drug", "fertilizer", "baking flour"). A newly acquired product is of the same quality, but has potentially harmful properties or consequences.

> (Vignette "Hospital")
> "Dr Jones works as doctor in a local hospital. Dr Jones often administers her patients the blood-thinning drug "Heparine" in order to prevent thrombosis and blood clots. Normally, blood-thinning drugs do not cause any side effects with certain blood types.
>
> The hospital has recently started to order an additional blood-thinning drug, 'Afibo', that is cheaper than 'Heparine'. 'Afibo' is as effective as 'Heparine', but has one side effect. It causes mild leg cramps in patients with blood type 'AB-positive'. "

In dependence on the 'knowledge' condition, the middle part of the vignette manipulated whether the agent possesses relevant knowledge about the harmful properties of the item.

> *Knowledge / No Knowledge* "Although [Because] the drug 'Afibo' has only recently been ordered, Dr Jones knows [does not know] that this drug causes mild leg cramps in patients with blood type 'B-negative'. "

After reading the first part of the vignette, participants had to answer two comprehension check questions. Participants then proceeded to the last part of the vignette. The final part of the vignette described the agent's (knowing or unwitting) use of the item, resulting in harmful consequences.

Causal Question After the final part of the vignette, participants had to answer a causal rating question, and generate a counterfactual alternative in an open-text response. The causal rating question asked participants to what extent they agree with the statement "Dr Jones [agent] caused the patient's leg cramps [outcome]" on a 7-point Likert scales (1-'strongly disagree', 7-'strongly agree').

Counterfactual Question In order to probe counterfactual thoughts, participants were instructed to write down what could have gone differently so that the patient would not have suffered mild leg cramps in a free text response ("If _____, the patient would not have suffered leg cramps [effect absent]"). This open-text counterfactual question allowed us to elicit the individual point of intervention in people's imagined alternative scenarios. Based on participants' written responses, we developed a four-part coding rubric. The first category "Action" (N = 112) covered all responses that described a change of *just* the agent's action that led to the outcome, i.e. the use of the item that caused the outcome (fertilizer/drug/baking flour) (e.g. *"If Dr Smith had not administered Corus to the patient"*). The second category "Epistemic state" (N = 75) covered all responses that described a change in the agent's epistemic states (e.g. *"If Dr. Jones had known about the side effects of Afibo..."*, unspecified how / caused by the agent / caused by others). Remaining answers did not show a specific theme, so we clustered the answers around two broad categories. The third category included any changes related to the causal agent, "Agent-related" (N = 30) (*additional action / prior action / character trait*). The fourth category "Environment" (N = 25) included all kinds of changes that did not relate to the agent (*change in environment / modifications in the properties of the item used, etc.*). Participants' responses were coded by the first author and a researcher assistant. Inconsistent codes were resolved by discussion.

Results

A likelihood ratio test indicated that a model including knowledge provided a better fit for the people's causal ratings than a model without it, $\chi^2(1) = 92.5$; $p < .001$. People's causal ratings decreased (b = -2.03, SE = .35, t = -5.87) when the agent was ignorant (M =3.96, SD = 2.20, 95% CI [3.57, 4.35]) compared to a knowing agent (M =6.22, SD = 1.30 95%, CI [5.99, 6.46]) (see Figure 1). Adding scenario ($\chi^2(2)$ = 2.66; p = .27) and an interaction with knowledge ($\chi^2(2)$ = 5.37; p = .07) did not provide a significantly better fit to the data. A multinomial logistic regression was performed to model the relationship between the knowledge condition and the type of counterfactual response ("action", "epistemic state", "agent-related", "environment"), with "environment" as reference category. Addition of the knowledge predictor to a model that contained only the intercept significantly improved the fit between model and data, $\chi^2(3)$ = 121.49; $p <$ .001, R^2= .21. When the agent's epistemic state changes from knowledge to ignorance, people are less likely to imagine a counterfactual change that concerns the agent's action (b = -1.46, OR = .23, SE = .46, z = 3.16, $p < .001$), more likely to imagine a change in the agent's epistemic state (b = 2.77, OR = 16.00, SE = .72, z = 3.86, $p < .001$) and less likely to indicate an agent-related change (b = -2.01, OR = .13, SE = .64, z = -3.16, $p < .01$), compared to a change in the environment. (see Figure 2).

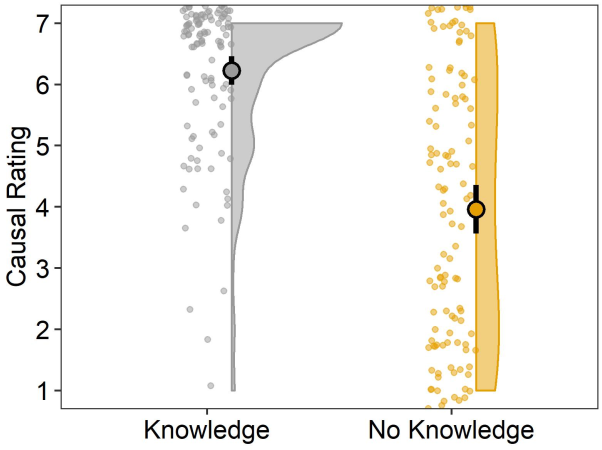

Figure 1: **Experiment 1** Causal Ratings. Big dots are group means , error bars depict 95% Confidence Intervals. Coloured backgrounds represent the probability distribution of the data, small dots are individual participants' judgments.

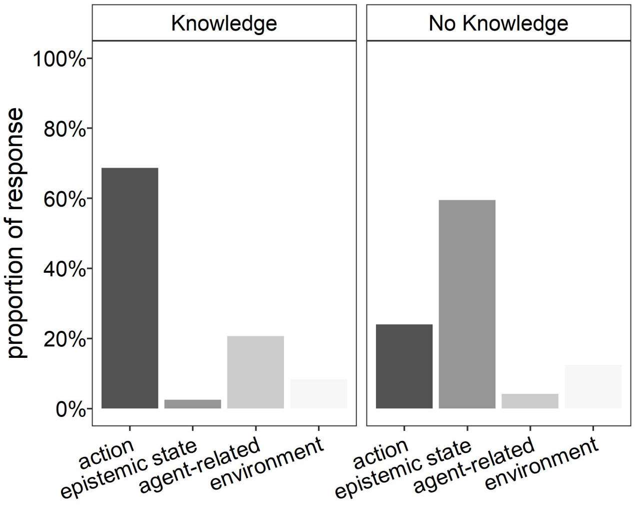

Figure 2: **Experiment 1**: Counterfactual responses

Discussion

The first experiment replicated previous findings demonstrating the influence of agentic epistemic states on people's causal attributions (Lagnado & Channon, 2008; Lombrozo, 2010; Gilbert et al., 2015). Ignorant agents are perceived as less causal than knowledgeable agents. At the same time, the agent's epistemic state also shifts the focus of an imagined counterfactual change in the actual scenario. In case of ignorance, people are less likely to refer to a change in the agent's causal action, but prefer to imagine a change in the agent's knowledge state. In the scenarios of Experiment 1, the reasons for the agent's ignorance about the outcome were underspecified. The vignette leaves open whether and to what extent the agent could have changed their own state of ignorance. We were therefore interested if the conditions under which an agent's ignorance came about also influence how causal the agent is perceived, as well the kind of counterfactuals people imagine.

Experiment 2

In the second experiment, we aimed to assess judgments about agents whose ignorance was either self- or externally caused.

Participants and Design

We recruited 179 participants on Amazon Turk. 27 participants were excluded for not answering all eight comprehension check questions correctly, and two participants were excluded for providing a nonsensical counterfactual responses. The final sample consisted of 150 participants (M_{age} = 37.78, SD_{age} = 11.67, N_{female} = 50). We adopted a 2 ignorance (self-caused vs. externally caused) × 3 scenario ("hospital" vs. "garden" vs "bakery") design. 'Ignorance' was manipulated within participants and 'scenario' was manipulated between participants.

Material

In Experiment 2, agents were ignorant about the consequences of their action in both conditions, and we manipulated how their state of ignorance was brought about. In this vignette, an e-mail that contains the relevant information about the harmful properties of an item is sent to the agent. In the "*externally caused ignorance*" condition, this e-mail is deleted due to a technical default. In the "*self-caused ignorance*" condition, the agent does not obtain the information because they fail to read the e-mail.

> *Externally caused* "Dr Jones checked her inbox, but she did not see the e-mail of the pharmacy manager because it was erroneously marked as spam and automatically deleted from the account."/ *Self-caused* "Dr Jones read her inbox and saw the e-mail of the pharmacy manager, but did not read it."

In both conditions, the scenarios ends with the agent applying the relevant item, unknowing about the harmful properties of the item. As a result, a bad effect obtains.

Causal Rating and Counterfactual Question Causal and Counterfactual Question were asked as in Experiment 1. We excluded the responses from eight participants who indicated that the agent in the "externally caused ignorance" condition could have looked into the spam-folder and read the e-mail, signalling a misunderstanding of the scenario. The first category "Action" (N = 11) was used as in Experiment 1. "Direct epistemic change" (N = 11) referred to responses that suggested a direct change of the agent's knowledge about the item without specifying how (*"If Sandra had known about the walnuts..."*). "Epistemic state change by agent action" (N = 173) included all types of epistemic state changes of which the agent was the primary cause (*by reading the e-mail / by an additional information-gathering action*, etc.). "Epistemic state change by other" (N = 92) included changes in the agent's epistemic that were not primarily caused by an action of the agent themselves (*by information given by a third-party agent, by a change in the e-mail system etc.*). Finally, the category "Environment" (N = 5) included changes in the environment or setting that did not affect the agent's causal action or epistemic state.

Knowledge and Blame for Ignorance In addition to causal and counterfactual questions, participants had to indicate their agreement with the statement "Dr Jones [agent] could have known that 'Afibo' causes leg cramps [effect]" on a 7-point Likert scale (1-'strongly disagree', 7-'strongly agree'). This question allowed us to examine the degree to which people perceived the agent's ignorance as mutable. Finally, participants answered the question "How blameworthy is Dr Jones [agent] for not knowing that 'Afibo' causes leg cramps [effect]?" on 7-point agreement scale (1-'Not at all', 7-'Completely'). In addition to the perceived possibility of knowledge, we also wanted to assess people's judgments about the agent's blameworthiness for their ignorance.

Results

Causal and Counterfactual Question A Likelihood ratio test indicated that type of ignorance was a significant factor in predicting participant's causal responses, $\chi^2(1) = 108.54$; $p <$ .001. People's causal ratings decreased (b = -2.21, SE = .29, t = -7.69) when the agent's ignorance was caused externally (M = 3.52, SD = 2.19, 95% CI [3.17, 3.87]) rather than by choice (M = 5.73, SD = 1.59 95%, CI [5.48, 5.98]) (see Figure 3). There was no significant effect of scenario (p = .90) and no interaction between ignorance and scenario (p = .99).

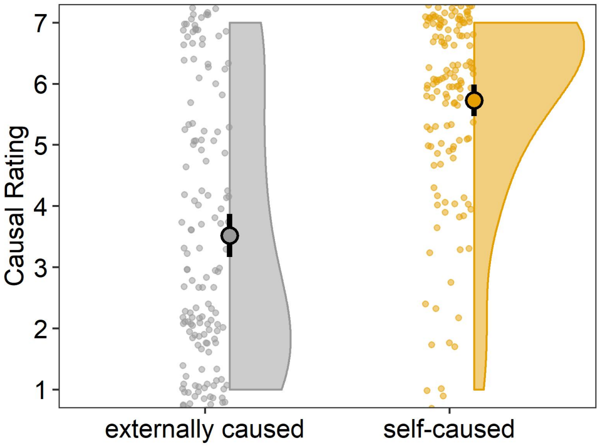

Figure 3: **Experiment 2**: Causal Ratings.

Addition of the ignorance predictor to a multinomial logistic regression model that contained only the intercept significantly improved the fit for predicting counterfactual responses, $\chi^2(3) = 146.41$; $p < .001$, R^2= .26. Changing the epistemic condition of ignorance from self-caused to externally caused is associated with a decrease in indicating a self-caused epistemic change (b = -1.39, OR = .24, SE = .63, z = -2.19, p = .03), and an increase in an externally caused epistemic change responses (b = 2.93, OR = 16.00, SE = .79, z = 3.67, $p < .001$) (see Figure 4).

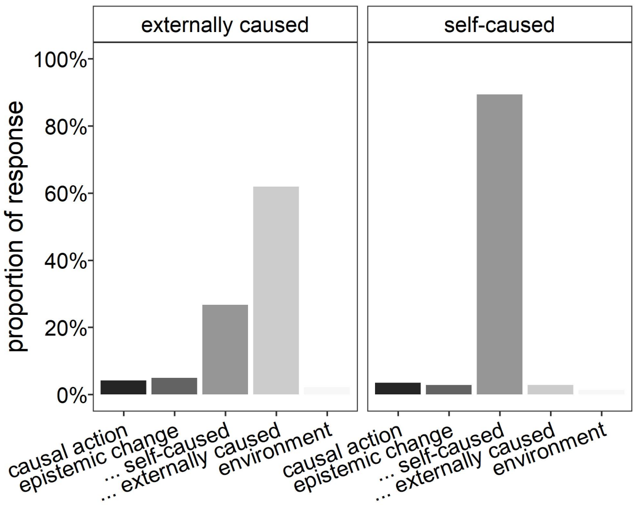

Figure 4: **Experiment 2**: Counterfactual responses.

Knowledge and Blame The condition under which ignorance came about significantly predicted people's modal judgement about the agent's epistemic state, $\chi^2(1) = 114.50$; $p < .001$. People's agreed less that the agent could have known (b = -2.64, SE = .29, t = -8.40) when the agent was ignorant because of a technical default (M = 3.55, SD = 2.11, 95% CI [3.20, 3.90]) compared to ignorance caused by the agent themselves (M = 6.02, SD = 1.69 95%, CI [6.72, 5.77]). Type of ignorance also influences people's judgement about the agent's blameworthiness for their ignorance, $\chi^2(1) = 237.15$; $p < .001$ (b = -3.46, SE = .28, t = -12.15), with people assigning less blame when the ignorance was externally caused (M = 2.75, SD = 1.84, 95% CI [2.40, 3.10]) vs. self-caused (M = 6.09, SD = 1.27, 95% CI [5.84, 6.35]).

Discussion

Experiment 2 demonstrated that the epistemic condition of ignorance influences people's causal judgement about the agent, their modal judgements about the agent's epistemic state, as well as how blameworthy the agent is considered for their ignorance. The perceived causal and blame difference is mirrored by the target of counterfactual change. In dependence of whether the access to relevant information is prevented by an external cause or the agent's own actions, people vary in how likely they are to imagine an epistemic state that is brought about by the agent's action. Notably, a substantial proportion of people (26%) still indicated a self-caused epistemic change in the "externally caused ignorance" condition, mostly by referring to alternative information-seeking actions the agents could done. This finding suggests that people generally give weight to agentic actions when imagining how an agent's epistemic states could have been different. According to counterfactual theories of causation, causality is determined by the counterfactual dependence of the outcome on the candidate cause in the actual world, but also under different 'contingencies', e.g. when background circumstances are different (Chockler & Halpern, 2004; Halpern, 2016; Gerstenberg & Lagnado, 2014). In Experiment 3, we want to apply this idea to epistemic conditions. That is, we wanted to test whether people take into account agents' epistemic actions, even if these actions do *not* lead to the acquisition of knowledge in the actual scenario, but would have under different circumstances.

Experiment 3

In our third experiment, we were interested in testing whether people take into account an agent's information acquisition under actual and possible circumstances.

Participants and Design

We recruited 171 participants on Amazon Mechanical Turk. 34 participants were excluded for failing one or more of the four comprehension check questions, and 2 participants were excluded for providing a non-sensical counterfactual response, leaving a final sample size of $N = 133$ ($M_{\text{age}} = 38.36$, $SD_{\text{age}} = 11.38$, $N_{\text{female}} = 57$, 1 = unidentified). We adopted a 2 ignorance (information search vs. no information search) × 3 scenario ("hospital" vs. "garden" vs. "bakery") design. 'Information acquisition' was manipulated within participants and 'scenario' was manipulated between participants.

Material

In the frame story of Experiment 3, an email about the relevant item is (successfully) sent to the agent. However, in this e-mail, the crucial information about the harmful property of the item is missing. We then varied whether the agent read ("information-seeking") or did not read the e-mail ("not information-seeking"). As before in Experiment in 2, in both conditions the agent unwittingly applies the harmful item with negative consequences.

Rating Questions & Counterfactual Question Causal Ratings, Counterfactual Question as well as Knowledge and Blame Ratings were asked as in Experiment 2. We excluded those counterfactual categories from the analyses that had less than 5% of participants' responses across both "information-seeking" conditions. "Epistemic state" (N = 17) included all responses that stated an epistemic change without indicating how, "... by info" (N = 90) referred to responses indicating the presence of the relevant information in the e-mail and "by info + reading e-mail" (N = 44) added the action of reading the e-mail to the response. The category "by additional action of agent" (N = 61) referred to responses indicating the agent acquiring knowledge by additional means, and "... by someone else" (N = 24) encompassed all responses that indicated an epistemic state change induced in the agent by a third-party agent.

Forward-looking causal judgments In order to investigate whether people's causal judgments about the actual scenario reflect their causal considerations under different contingencies, we included a follow-up scenario. Participants were prompted to imagine a future scenario in which there is a new pain killer "Innohep" (bakery: flour brand, garden: weed killer) in hospitals. However, this pain killer causes nausea in patients who take beta-blockers. As usual, an e-mail has been sent out to all doctors that introduces the new pain killer, but this time the e-mail includes the information that this pain killer causes nausea in patients taking beta-blockers. Participants were then asked to estimate the likelihood that the agent from the "information-seeking" and the agent "non-information seeking" would read that e-mail in this future scenario: "How likely is it that Dr Jones [Dr Smith] would check the e-mail of the pharmacy manager about 'Innohep'?" (0 - "Extremely unlikely"; 100 - "Extremely likely"). In addition, they were asked about the likelihood of a bad outcome given that either agent would be in charge of a patient with the sensitive condition: "How likely is it that a patient who takes beta-blockers would suffer from nausea if Dr Jones were treating this patient [Dr Smith were treating this patient]" (0 - "Extremely unlikely"; 100 - "Extremely likely"). The responses to these questions serve as a proxy for people's forward simulation of the agents' future causality based on their prior epistemic actions.

Results

Causal and Counterfactual Question The "information seeking" factor was a significant predictor for participants' causal responses, $\chi^2(1) = 41.33$; $p < .001$. People judged the agent to be less of a cause (b = -.71, SE = .22, t = -3.29) when the agent read the e-mail with the missing information (M = 3.07, SD = 2.22, 95% CI [2.69, 3.45]) than if they did not (M =3.96, SD = 2.14, 95%, CI [3.6, 4.3]) (see Figure 5). The information acquisition condition also significantly predicted people's counterfactual responses $\chi^2(4) = 140.73$; $p < .001$, R^2= .21. When the agent did not read the e-mail, people were less likely to indicate a change that consisted in the addition of *just* the missing information in the e-mail (b = -2.59, OR = .07, SE = .63, z = -4.14, $p < .001$), compared to a change in just the epistemic state (see Figure 6).

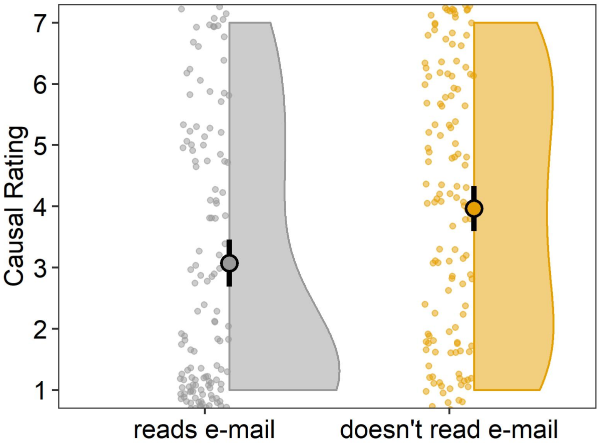

Figure 5: **Experiment 3**: Causal Ratings.

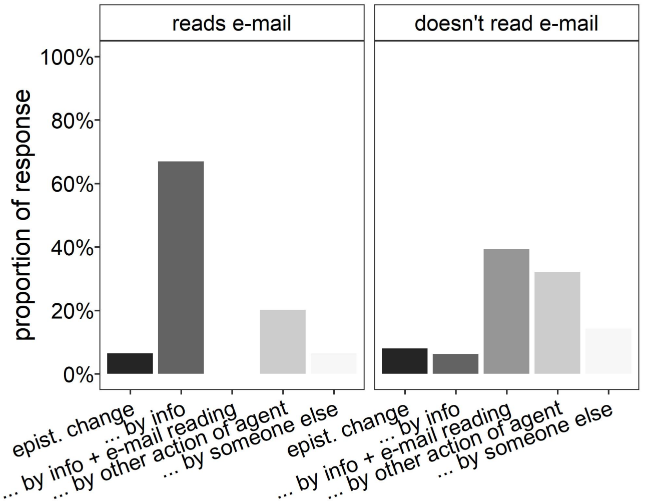

Figure 6: **Experiment 3**: Counterfactual Responses.

Knowledge and Blame Rating Information-seeking behaviour significantly predicted modal judgments about the agent's epistemic state $\chi^2(1) = 14.08$; $p < .001$, as well as blameworthiness for ignorance $\chi^2(1) = 55.56$; $p < .001$. The agent who did not read the e-mail containing missing information was judged to could have known about the relevant information to a slightly greater extent (M = 3.42, SD = 2.24, 95% CI [3.05, 3.78]) and to blame slightly more for their ignorance (M = 3.47, SD = 2.13, 95% CI [3.10, 3.84]) than the information-seeking agent ("Could have known": M = 2.90, SD = 2.24, 95% CI [2.53, 3.29]; "Blame": M = 2.43, SD = 2.04, 95% CI [2.05, 2.81]).

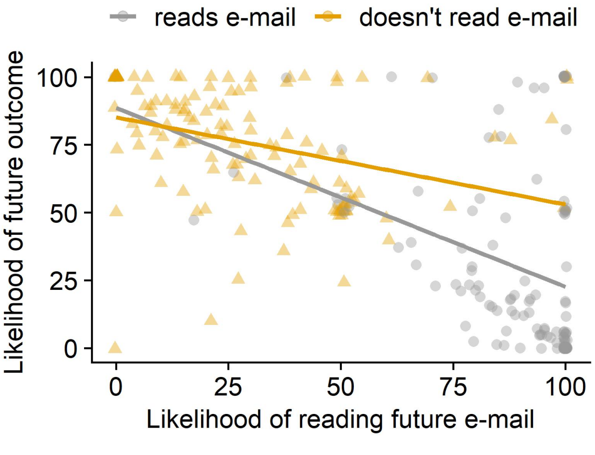

Figure 7: **Experiment 3**: Likelihood of agent reading a e-mail (X-Axis) and likelihood of future outcome (Y-Axis), grouped by previous epistemic action (reading vs. not reading e-mail).

Forward looking causation The likelihood of a bad outcome in a future scenario was predicted by the likelihood of the agent's information acquisition $\chi^2(1) = 38.05$; $p < .001$

as well as an interaction between the e-mail reading condition and the likelihood of information acquisition $\chi^2(1) = 4.16$; $p = .041$. Likelihood of outcome was negatively predicted by likelihood of information acquisition when the agent had read the e-mail (b = -.66, SE = .16, t = -4.18), and to a weaker extent when the agent hadn't read the e-mail (b = -.32, SE = .07, t = -4.30) (see Figure 7).

Discussion

Experiment 3 showed that an agent who unsuccessfully attempts to acquire knowledge is still seen as less causal for the unforeseen outcome of their action than an agent who does not attempt to do so, even if the attempt would be have been equally unsuccessful. We also found this difference in people's judgments about blame as well as their judgments about whether the agent *could* have known about the outcome. The fact that information-seeking is taken into account for the perceived causal strength of the agent likely results from people integrating alternative scenarios with different circumstances into their counterfactual thinking. In a world in which the e-mail had contained the relevant information, the agent who read the e-mail would have found out about the negative outcome, and the outcome would potentially not have occurred. People's forward-looking causal judgments in the follow-up scenario supported this.

General Discussion

In three studies, we have investigated people's causal and counterfactual reasoning about ignorant agents. We have shown that people judge an agent as less causal if i) they are ignorant vs. knowledgeable about their action consequences, ii) their ignorance was externally vs. self-caused and iii) their epistemic action in the actual scenario would have made a difference to their epistemic state under different circumstances. Crucially, these differences in causal judgments were mirrored in respective responses about counterfactual changes in agents' epistemic states or epistemic actions. Our results support the idea that people use a causal model that includes different levels of epistemic states and epistemic actions, and that people use counterfactuals over these models to assign causality to agents. Likewise, our results show that our manipulations also affected judgments about blame for ignorance. Causal judgments therefore could have been affected by how blameworthy people judged the agent for their ignorance, and in consequence by their judgements of blame for the unforeseen outcome (Alicke & Rose, 2012). While further research is needed to address the exact relationship between cause, blame and counterfactuals in these scenarios, this study lays the basis for the fundamental role of epistemic states in causal and counterfactual reasoning.

References

Alicke, M. D., & Rose, D. (2012). Culpable control and causal deviance. *Journal of Personality and Social Psychology Compass*, *6*, 723–725.

Alicke, M. D., Rose, D., & Bloom, D. (2012). Causation, norm violation, and culpable control. *The Journal of Philosophy*, *108*(12), 670–696.

Chockler, H., & Halpern, J. Y. (2004). Responsibility and blame: A structural-model approach. *Journal of Artificial Intelligence Research*, *22*, 93–115.

Gerstenberg, T., Goodman, N. D., Lagnado, D., & Tenenbaum, J. (2020, Mar). *A counterfactual simulation model of causal judgments for physical events.* PsyArXiv. doi: 10.31234/osf.io/7zj94

Gerstenberg, T., & Lagnado, D. A. (2014). Attributing responsibility: Actual and counterfactual worlds. In J. Knobe, T. Lombrozo, & S. Nichols (Eds.), *Oxford studies in experimental philosophy* (Vol. 1, pp. 91–130). Oxford University Press.

Gilbert, E. A., Tenney, E. R., Holland, C. R., & Spellman, B. A. (2015). Counterfactuals, control, and causation: Why knowledgeable people get blamed more. *Personality and Social Psychology Bulletin*, *41*(5), 643–658.

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

Hilton, D. J., McClure, J., & Moir, B. (2016). Acting knowingly: effects of the agent's awareness of an opportunity on causal attributions. *Thinking & Reasoning*, *22*(4), 461–494.

Kirfel, L., & Lagnado, D. (2021). Causal judgments about atypical actions are influenced by agents' epistemic states. *Cognition*, *212*, 104721.

Lagnado, D. A., & Channon, S. (2008). Judgments of cause and blame: The effects of intentionality and foreseeability. *Cognition*, *108*(3), 754–770.

Lewis, D. (2013). *Counterfactuals*. John Wiley & Sons.

Lombrozo, T. (2010). Causal–explanatory pluralism: How intentions, functions, and mechanisms influence causal ascriptions. *Cognitive psychology*, *61*(4), 303–332.

McClure, J., Hilton, D. J., & Sutton, R. M. (2007). Judgments of voluntary and physical causes in causal chains: Probabilistic and social functionalist criteria for attributions. *European journal of social psychology*, *37*(5), 879–901.

McGill, A. L., & Klein, J. G. (1993). Contrastive and counterfactual reasoning in causal judgment. *Journal of Personality and Social Psychology*, *64*(6), 897.

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

Samland, J., & Waldmann, M. R. (2016). How prescriptive norms influence causal inferences. *Cognition*, *156*, 164–176.

Schaffer, J. (2005). Contrastive causation. *The Philosophical Review*, *114*(3), 327–358.

Spellman, B. A., & Gilbert, E. A. (2014). Blame, cause, and counterfactuals: The inextricable link. *Psychological Inquiry*, *25*(2), 245–250.

Woodward, J. (2007). Interventionist theories of causation in psychological perspective. *Causal learning: Psychology, philosophy, and computation*, 19–36.