Cognition 218 (2022) 104924

Contents lists available at ScienceDirect

Cognition

journal homepage: www.elsevier.com/locate/cognit

Original Articles

The role of mechanism knowledge in singular causation judgments☆

Simon Stephan[*], Michael R. Waldmann

University of Göttingen, Germany

A R T I C L E I N F O

Keywords:
Singular causation
Causal attribution
Mechanisms
Causal Bayes nets
Causal chains
Computational modeling

A B S T R A C T

Singular causation queries (e.g., "Did Mary's taking contraceptives cause her thrombosis?") are ubiquitous in everyday life and crucial in many professional disciplines, such as medicine or law. Knowledge about general causal regularities is necessary but not sufficient for establishing a singular causation relation because it is possible that co-occurrences consistent with known regularities are in an individual case still just coincidental. Thus, further cues are helpful to establish a singular causation relation. In the present research we focus on information about mechanisms as a potent cue. While previous studies have shown that reasoners consider mechanism information as important when it comes to answering singular causation queries, no formal model has been proposed that explains why this is case. We here present a computational model that explains how causal mechanism information affects singular causation judgments. We also use the model to identify conditions that restrict the utility of mechanism information. We report three experiments testing the implications of our formal analysis. In Experiment 1 we found that reasoners systematically use mechanism information, largely in accordance with our formal model, although we also discovered that some people seem to rely on simpler, computationally less demanding reasoning strategies. The results of Experiments 2 and 3 demonstrate that reasoners have a tentative understanding of the conditions that restrict the utility of causal mechanism information.

1. Introduction

The main focus of past research on causal reasoning has been on how we acquire and use knowledge about causal regularities (e.g., "smoking causes lung disease"). It has been shown that such knowledge can support various causal inference types, including predictions, diagnoses, or explanations (see Sloman, 2005; Waldmann, 2017, for overviews). In the present research we investigate how reasoners determine the singular cause(s) of an observed event. As opposed to general causation queries that focus on general causal regularities (e.g., "Do contraceptives cause thrombosis?"), singular causation queries refer to causal connections between events that actually occurred in a particular place at a particular time (e.g., "Was Mary's thrombosis caused by the contraceptives she took?") (see Danks, 2017; Russo & Williamson, 2011, for overviews). Even when a general probabilistic causal relationship between two types of events has been established, the co-occurrence of the two events may still be a coincidence. How do reasoners determine whether a potential cause c actually caused an observed effect e or whether the co-occurrence was a coincidence and e was *actually* caused by some alternative cause a? The question of how people answer singular causation queries is relevant not only because such queries are prevalent in our everyday lives, but because they are also frequently asked in a number of professional disciplines, such as medicine or law (cf. Hart & Honoré, 1985; Lagnado & Gerstenberg, 2017; Russo & Williamson, 2011). Given that causal connections between events are not directly observable, the question is how reasoners decide that a co-occurrence of events is actually causal.

A widespread view both in philosophy and psychology is that reasoners must use their general causal knowledge to answer singular causation queries (Cheng & Novick, 2005; Danks, 2017; Hitchcock, 2009; Lagnado & Gerstenberg, 2017; Lagnado et al., 2013; Stephan et al., 2018, 2020; Stephan & Waldmann, 2018). According to recent psychological computational models of singular causation judgments (Cheng & Novick, 2005; Stephan et al., 2020; Stephan & Waldmann, 2018), one type of general causal knowledge crucial for the assessment of singular causation is knowledge about the strength of the potential causes, which can be induced from observable patterns of statistical dependencies (Cartwright, 1989; Cheng, 1997; Cheng & Lu, 2017;

☆ The materials and data of the reported studies can be accessed at: https://osf.io/325pr/?view_only=b6bbd6f423104997ac04ead54b0cd32d.

[*] Corresponding author at: Department of Psychology, University of Göttingen, Gosslerstr. 14, 37073 Göttingen, Germany.
E-mail address: simon.stephan@psych.uni-goettingen.de (S. Stephan).

https://doi.org/10.1016/j.cognition.2021.104924
Received 14 November 2020; Received in revised form 28 September 2021; Accepted 29 September 2021
Available online 19 October 2021

Griffiths & Tenenbaum, 2005; Novick & Cheng, 2004) or through analogical reasoning (Holyoak et al., 2010). Another relevant type of knowledge is temporal knowledge, such as knowledge about the causal latency of the potential causes (Stephan et al., 2018, 2020).

The present research focuses on a further cue that has been identified in the philosophical literature (Cartwright, 2015, 2017; Danks, 2005) as important for establishing singular causation relations – knowledge about causal mechanisms. In psychological studies, causal mechanism information has been found to play a role in causal reasoning in a number of different contexts (Ahn & Bailenson, 1996; Ahn et al., 1995; Hegarty, 2004; Johnson & Ahn, 2015; Lombrozo, 2010; Park & Sloman, 2013; see Johnson & Ahn, 2017, for a recent overview). Moreover, in line with different philosophical accounts, studies have shown that people, even young children (Buchanan & Sobel, 2011; Cimpian & Erickson, 2012), consider mechanism information to be particularly relevant for the assessment of singular causation (Ahn & Bailenson, 1996; Ahn et al., 1995; Johnson & Keil, 2018). For example, in a classic study by Ahn et al. (1995) in which subjects were asked to come up with an explanation for, for example, why “John had an accident on Route 7 yesterday”, most subjects asked questions about the presence of a possible mechanism leading to accidents instead of questions about covariation information. In a more recent study by Johnson and Keil (2018) subjects were asked to evaluate both general and singular causation claims. This study found an interesting dissociation: While subjects who were asked to evaluate general causation claims (e.g., “Smoking causes cancer”) preferred to consult covariational information, subjects asked to evaluate singular causation claims (e.g., “Jack's smoking caused his cancer”) preferred information about causal mechanisms that could link the two events.

While previous studies have yielded interesting insights into causal reasoning and clearly documented that causal mechanism information seems to be considered by people as a relevant cue when it comes to the assessment of singular causation relations, what has been lacking in the causal reasoning literature is a formal theory that explains why reasoners seek for causal mechanism information when aiming to establish the actual singular cause(s) of an observed target effect. The goal of the present research is to fill this gap and to provide such a computational account. We focus on five key questions in this paper: (1) Why is causal mechanism information useful for the assessment of singular causation? (2) How can causal mechanism information be incorporated into a formal computational model of singular causation judgments? (3) How well does the model explain lay people's singular causation judgments and to which extent are people's singular judgments sensitive to the different relevant factors identified by the model? (4) What are the conditions that constrain how useful causal mechanism information is in the assessment of singular causation? (5) Do people recognize these constraints and incorporate them in their singular causation judgments?

The basis of our formal analysis is the power PC framework of causal attribution proposed by Cheng and Novick (2005), which was later developed further into the generalized power PC model of singular causation (Stephan et al., 2018, 2020; Stephan & Waldmann, 2018). We will start with a summary of this new model (Stephan et al., 2020; Stephan & Waldmann, 2018), and then propose an extension that incorporates causal mechanism information. The extended model not only formalizes under which circumstances mechanism information is helpful for the assessment of singular causation, it also allows us to identify situations in which mechanism information is less helpful. We will then present the results of three experiments in which we systematically tested the predictions of the new model.

2. The generalized causal power PC model of singular causation judgments

The generalized power PC model of singular causation judgments (Stephan et al., 2020; Stephan & Waldmann, 2018) extends Cheng and Novick's (2005) power PC model of causal attribution, which applies Cheng's (1997) causal power PC theory (see also Cheng & Buehner, 2012; Cheng & Lu, 2017; Liljeholm & Cheng, 2007) to situations in which a reasoner tries to decide whether an observed effect e is caused by a target cause c or an alternative cause a. The model can be illustrated using causal Bayes nets (Glymour, 2001; Gopnik et al., 2004; Pearl, 1988, 2000). The causal model in Fig. 2a describes a general causal relationship in which C and A are two independent generative causes of a common effect E. C and A are assumed to combine their influence according to a *noisy-OR* gate (Glymour, 2003; Griffiths & Tenenbaum, 2005; Meder et al., 2014; Pearl, 1988), which means that they generate the effect disjunctively with independent probabilities. The link parameters w_c and w_a represent these probabilities and are called the causal strengths of the causes (Cheng, 1997; Griffiths & Tenenbaum, 2005, 2009). The noisy-OR parametrization implies that when only C occurs, it generates E with probability w_c, i.e., $P(e|c,\neg a,w_c) = w_c$, when only A occurs it generates E with probability w_a, i.e., $P(e|a,\neg a,w_a) = w_a$, and when both occur they have independent chances to produce E, i.e., $P(e|c,a,w_c,w_a) = w_c + w_a - w_c \cdot w_a$ (see also Griffiths & Tenenbaum, 2005, p. 346).

To illustrate how the generalized power PC model of singular causation judgments works, we assume a singular case in which both C and E are present (i.e., $C = 1$ or c, $E = 1$ or e) and a reasoner wonders whether e was actually caused by c. We assume that the alternative cause A is also present (i.e., $A = 1$ or a). According to the model, the probability that c caused e in this case is given by:

$$P(c \rightarrow e|c,a,e) = \frac{w_c - w_c \cdot w_a \cdot \alpha}{w_c + w_a - w_c \cdot w_a} = \frac{w_c \cdot (1 - w_a \cdot \alpha)}{P(e|c,a)}. \quad (1)$$

We first neglect the model's α parameter by assuming that its value is 0 in the current situation. The numerator reduces to the target cause's causal strength w_c in this case, which is then normalized by the conditional probability of the effect given C and A. The model determines the relative frequency of cases among all co-occurrences of C and E in which C was strong enough to generate E, taking into consideration that A can also sometimes generate E. The model's predictions for different parameter value combinations are shown in Fig. 1. The darkest line in each panel captures the type of situation we are focusing on in the moment, situations in which $\alpha = 0$. The graphs show that the model captures two prima facie reasonable ways of reasoning about singular causation. First, it predicts that reasoners should be more confident that c caused e, the stronger C generally is (i.e., the higher w_c is). Second, it predicts that reasoners should be increasingly confident that c caused e, the weaker the alternative cause(s) A is (i.e., the lower w_a is). If $w_a = 0$, the model predicts that c must have caused e, no matter how weak C generally is. The model thus captures a way of reasoning about singular causation that corresponds to what has been called “Holmesian inference” or “reasoning by elimination of alternatives” (Bird, 2005, 2007, 2010) – the inferential process by which one of multiple mutually exclusive explanations for a phenomenon is selected based on the relative probabilities of the competing explanations (see also Lipton, 2004; Lombrozo & Vasilyeva, 2017, for related accounts on abduction or *inference to the best explanation*).

We now explain the function of the α parameter. The product $w_c \cdot w_a \cdot \alpha$ in the equation's numerator accounts for the possibility of causal preemption of the target cause C by the alternative cause A. Causal preemption occurs when two potential causes are simultaneously sufficiently strong to generate the effect but only one of them actually succeeds in causing the effect (see Halpern & Pearl, 2005; Hitchcock, 2007, 2009; Paul & Hall, 2013). A famous example in the philosophical literature is a scenario involving two rock throwers, Billy and Suzy. Billy and Suzy are perfectly accurate rock throwers and neither of them, when acting alone, ever fails to hit and destroy a bottle (i.e., their causal strengths are 1.0). On a singular occasion both are aiming for the same bottle and both are throwing their rocks with identical speed. Suzy, however, manages to throw her rock a little bit earlier than Billy, and the bottle shatters. It is intuitively Suzy's and not Billy's rock throwing that

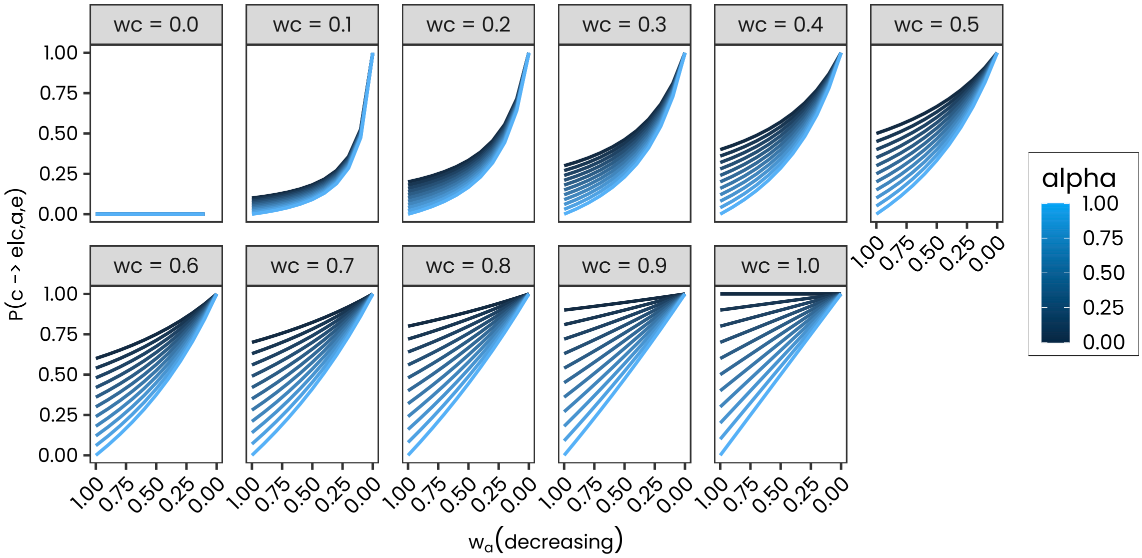

Fig. 1. Predictions of the generalized power PC model of singular causation judgments (Equation (1)) for different values of w_c, w_a, and α.

caused the bottle's breaking, even though we know that Billy's throwings are generally also successful. Situations in which the competing causes are simultaneously sufficient are identified by $w_c{\cdot}w_a$ in the equation's numerator. α represents a weighting parameter determining the proportion of the "sufficiency overlap" in which C was preempted by its competitor A. α has to be determined based on information about the temporal relation between the causes (Stephan et al., 2020). Relevant temporal factors determining α are the causal latencies (put briefly: the quicker a cause exerts its influence compared to its competitor, the less likely it is preempted by that competitor) and the onset difference between them (put briefly: a cause occurring earlier than its competitor is less likely to be preempted by it). Since $w_c{\cdot}w_a{\cdot}\alpha$ is the probability that C is preempted by A, it needs to be subtracted from w_c. α can take on any value between 0 and 1.0 (see Stephan et al., 2020). For example, $\alpha = 1$ models situations in which A definitely preempts C if both are simultaneously strong enough to generate E. An α value of 0.5 expresses uncertainty about the causes' preemptive relation. An α value of 0 models situations in which C is definitely not preempted by A.

Equation (1) applies to situations in which C and A are both present (i.e., to situations in which we conditionalize on C and A's presence). Situations in which one or both potential causes are unobserved can also be modeled. In this case the causes' strength parameters need to be multiplied with their respective base rate parameters, b_c and b_a.

3. Incorporating causal mechanism knowledge

We will now extend the model to address the question of how causal mechanism information can be incorporated into the judgment process. The general idea captured by the extended model is that causal mechanism information helps in the assessment of singular causation because it allows reasoners to insert more specific values for the different parameters of the original model.

The generalized power PC model of singular causation judgments was originally developed for a simple common-effect causal model in which C and A represent *direct* causes of the target effect E (see Stephan et al., 2020; Stephan & Waldmann, 2018). However, causal models express a reasoner's current state of causal knowledge, and they can be extended if additional knowledge about the mechanisms linking causes and effects becomes available, thereby turning a formerly direct causal connection into an indirect chain (see also Stephan et al., 2021). For example, we might first use a causal model in which the taking of Aspirin and relief from headaches are directly causally connected, $A \rightarrow H$. Later, we may learn about prostaglandin synthesis (P) as the underlying causal mechanism. This new mechanism knowledge can be incorporated into the causal model by changing the direct causal relation ($A \rightarrow H$) into an indirect one in which variable P serves as a mediator, $A \rightarrow P \rightarrow H$. Under the causal Bayes net framework, causal mechanisms are understood as more elaborate representations of sequences of causal dependencies (see, e.g., Johnson & Ahn, 2015; Stephan et al., 2021; Woodward, 2011).

Fig. 2A-C shows causal models with increasingly complex mechanism representations. All models represent the same general causal relation with two binary causes C ($C(\Omega) = 0, 1$, where $1 =$ "cause present" and $0 =$ "absent) and A ($A(\Omega) = 0, 1$) and one effect E ($E(\Omega) = 0, 1$), but they differ in terms of how much additional mechanism knowledge is represented. In Fig. 2B two intermediate binary nodes, M_C and M_A, connecting C and A to E are added. Here the mechanisms are modeled as simple causal chains in which the causal Markov condition holds, which states that, conditioned upon its direct causes, each variable in a causal network is independent of all other variables in the network except for its own direct and indirect effects. In the case of a simple causal chain $A \rightarrow B \rightarrow C$, for example, this means that given $B = 1$, the probability of $C = 1$ is invariant under $A = 1$ and $A = 0$. Using causal models augmented by intermediate nodes that represent mechanism variables, an important assumption that we make is that the initial (or global) causal strength parameters w_c and w_a of the causal model shown in Fig. 2A are now divided into two sub-components. For instance, C's initial strength parameter w_c is partitioned into w_{cmc}, the strength with which C generates its mechanism component M_C, and w_{mce}, the strength with which M_C generates E, and it holds that $w_c = w_{\text{cmc}}{\cdot}w_{\text{mce}}$ (cf. Waldmann et al., 2008a).

3.1. Observing that mechanisms were inactive in a singular case

Having specified how causal mechanism knowledge can be represented with causal models, we now show how knowledge about mechanism variables can help with determining the singular cause of an effect. To better illustrate how our model works, we will use our experimental scenario as a running example. An illustration of that scenario is shown in Fig. 3. In our studies, subjects learned about the medical emergency system of a medieval kingdom called Tristonia. The scenario description first introduced the King's palace in the North (the effect variable E) and two castles (the two candidate causes C and A), one located in the Southwest and the other in the Southeast of the empire. It was pointed out that the empire's only healer is living in the King's palace and that in case of medical emergencies the Western and Eastern castles must send out carrier pigeons (representing the cause

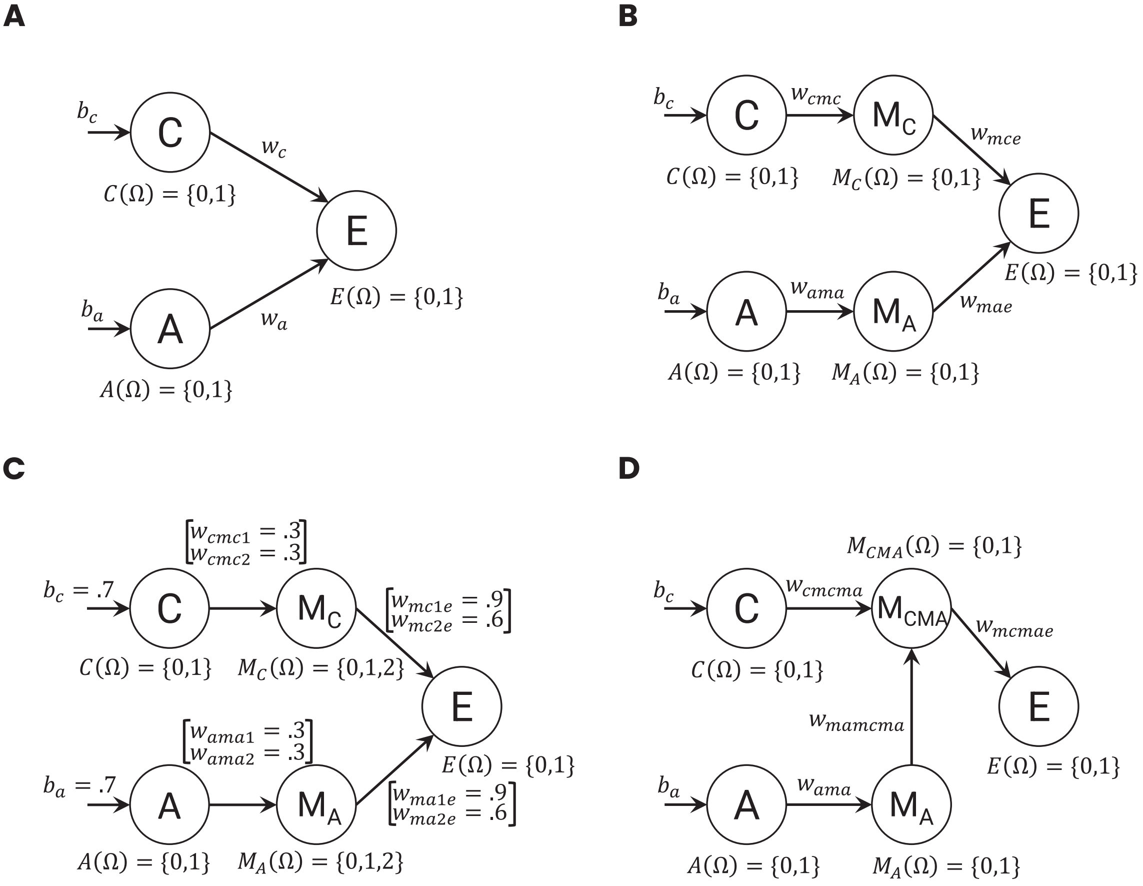

Fig. 2. Different causal models in which C and A are root causes of E. *Note.* A: a common-effect model with two binary causes and a binary effect. B: an augmented version of the model with causal mechanism variables. C: the same model but C and A here generate E via different alternative mechanisms, represented by ternary mechanism nodes. Here, the different strengths of the root causes (e.g., w_{cmc1} and w_{cmc2}) refer to the different possible states of the mechanism nodes (e.g., $M_C = 1$ and $M_C = 2$). For example, w_{cmc1} is the strength with which C causes $M_C = 1$. Analogously, the different strength parameters of the mechanism nodes refer to their possible states. For example, $w_{\mathrm{mc1}e}$ is the strength with which $M_C = 1$ causes E. The parameter values shown in this structure are the ones we used to generate the model predictions for Experiment 1. D: a common-effect causal model with intersecting mechanisms.

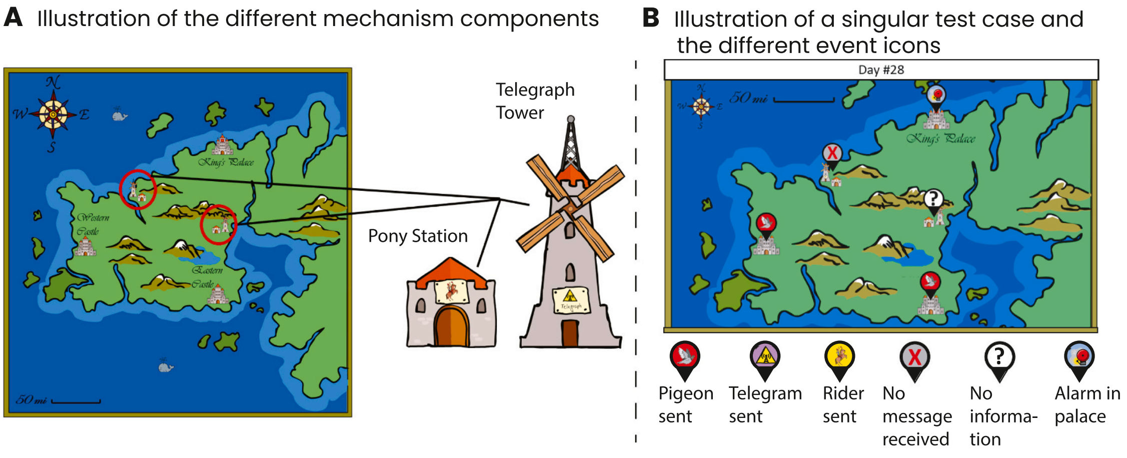

Fig. 3. Illustrations included in the instructions of Experiment 1.

events $C = 1$ or $A = 1$) to trigger an alarm in the King's palace (representing the effect event $E = 1$). It was emphasized that the pigeons cannot fly the whole distance to the palace, but can only reach intermediate stations located halfway between a castle and the palace (representing the mechanism nodes M_C and M_A). In Experiment 1 we introduced two alternative possibilities of how the emergency signals from the castles could be forwarded to the palace at these intermediate stations. One possibility were telegraphs and another were pony riders. The pony riders were introduced as back-ups for occasions on which a telegraph tower is blacked out. Below we will say more about these two different mechanism components. For the moment it is sufficient consider the special case that only one of these mechanism components exists, for example, telegraphs, and that the intermediate stations thus represent simple binary variables (e.g., $M_C = 1 =$ telegraph sent vs.

$M_C = 0 =$ no telegraph sent). In our experiments, subjects were shown different pictures that showed what happened on different days. The different possible events were symbolized by little event icons displayed above the different components. An illustration of a test case is shown in Fig. 3B. In all test cases, subjects saw that an alarm occurred in the King's palace, while the states of all the other nodes of the network were varied. For each test case, subjects were asked to indicate how strongly they believed that the alarm in the King's palace was caused by the Western [Eastern] castle.

To illustrate our model, we focus again on different situations in which a reasoner has observed that C, E, and an alternative cause A all co-occur, while the status of known intermediate mechanism variables varies. One possible observation a reasoner may make is that a known causal mechanism variable is inactive in the given situation. Consider the test case of our scenario shown in Fig. 3B. An alarm occurred in the King's palace on that day and both castles sent out a pigeon. We assume that the target cause is the Western castle. The status of the alternative cause's mechanism variable (i.e., the Eastern intermediate station) is unknown, but it is observed that the target cause's mechanism variable (i.e., the Western intermediate station) failed to become active on that day. In this case, the Western castle should be ruled out as the singular cause of the effect. This example illustrates that observing that mechanism variables are inactive in a singular case is helpful because it helps to identify the actual cause of the effect through the elimination of possible causes.

Importantly, our model formally captures this process. Just as w_c can be rewritten as $w_{\text{cmc}} \cdot w_{\text{mce}}$, $P(c \to e|c, a, e)$ can be rewritten as $P(c \to m_C|c, a, e) \cdot P(m_C \to e|c, a, e)$: The probability that c caused e corresponds to the probability that c caused m_C times the probability that m_C caused e (assuming the Markov condition holds). When we learn that the target cause's mechanism component was inactive in the given situation (i.e., $M_C = 0$ or $\neg m_C$), $P(c \to m_C|c, a, e)$ would be 0. Since $P(c \to e|c, a, e) = P(c \to m_C|c, a, e) \cdot P(m_C \to e|c, a, e)$, it follows that $P(c \to e|c, a, e)$ would also be 0. Discovering that the target cause's mechanism variable is inactive thus should lead us to rule out the target cause as the singular cause of e. Another way to look at this case is to consider what value w_c would take on if we conditionalized on $M_C = 0$. w_c would be 0 and, as a result, Equation (1) would also yield a value of 0.

What happens with $P(c \to e|c, a, e)$ in a situation in which a reasoner learns that the alternative cause's mechanism M_A was inactive? Consider the same test case as before but assume that the target cause C now is the Eastern instead of the Western castle. In this case, the probability that M_A caused e, $P(a \to m_A|c, a, e)$, would be 0. The probability that a caused e, $P(a \to e|c, a, e)$, would thus also be 0 because $P(a \to e|c, a, e) = P(a \to m_A|c, a, e) \cdot P(m_A \to e|c, a, e)$. Since the Western castle is the only possible alternative cause (A) of E, E is actually present (e) but a can be ruled out as its cause; here we should conclude that c must have caused e. This is formally captured by our model, as $P(c \to e|c, a, \neg m_A, e)$ is indeed 1.0 in this case. Conditionalizing on $M_A = 0$, A's influence on E is screened off, w_a takes on a value of 0, and Equation (1) reduces to $\frac{w_c}{w_c}$. Our model thus formally captures how learning about the inactivity of known mechanism variables supports an epistemic process that has been called eliminative or "Holmesian" reasoning (Bird, 2010).

3.2. *Observing that mechanisms were active in a singular case*

Next, we consider situations in which a reasoner learns that mechanism variables were active instead of inactive. Previous studies showed that reasoners often seem to search for present possible mechanism variables, and we demonstrate that observing that the target cause's mechanism is active should indeed make it more likely that this cause was the singular cause of e. Imagine a test case of our experimental scenario in which the Western castle is the target cause C and the Eastern castle is the alternative cause A. It is unknown whether A's intermediate station is active on that day (i.e., $M_A = ?$) but it is observed that C's intermediate station sent out a telegraph (i.e., M_C1). Observing that the intermediate station of the Western castle sent out a telegraph should make it more likely that the Western castle caused the alarm in the King's palace on that day.

Our model formally captures this intuition, as it holds that $P(c \to e|c, m_C, a, e)$ will on average be higher than the initial $P(c \to e|c, a, e)$. If C (the Western castle) is a necessary cause of its mechanism M_C, then $P(c \to m_C|c, m_C, a, e) = 1$, and $P(c \to e|c, a, e)$ therefore reduces to $P(m_C \to e|c, m_C, a, e)$. The reason why $P(m_C \to e|c, m_C, a, e)$ will on average be higher than $P(c \to e|c, a, e)$ is that the cause's global strength w_c (e.g., the strength with which sending out a pigeon causes an alarm in the palace) that enters into the calculation of $P(c \to e|c, a, e)$ corresponds to $w_{\text{cmc}} \cdot w_{\text{mce}}$ (e.g., the strength with which sending a pigeon activates the telegraph tower, and the strength with which the telegraphs lead to an alarm). This implies that w_{mce}, the strength parameter needed to compute $P(m_c \to e|c, m_C, a, e)$, will on average be higher than the global parameter w_c.

What if the status of M_C is unobserved but it is observed that the alternative cause's mechanism is active? $P(c \to e|c, a, e)$ becomes smaller in this case because the probability that a caused e simultaneously increases. w_A is substituted with w_{mae} in Equation (1). Since w_{mae} will on average be higher than w_a, observing $M_A = 1$ diminishes the numerator and increase the denominator of Equation (1). $P(e|c, a)$ in the denominator has to be replaced by $P(e|c, a, m_A) = P(e|c, m_A)$ in this case, which tends to be larger than $P(e|c, a)$.

Discovering that mechanism components were present in a given situation can also be helpful in other situations. We have so far considered situations in which a reasoner already knows that the potential causes of the effect were present (c, a). In many situations, however, reasoners will not have observed the potential causes of an effect e. In such cases, examining the status of mechanism variables mediating the influence of potential causes can help to reconstruct and narrow down the set of potential causes through a *diagnostic inference* about the likely causes of the observed mechanism variables. As an example, consider the case of an autopsy that reveals high concentrations of a particular substance Y in the victim's blood that typically occur after a person ingested a certain type of poison X. If the coroner knows that poison X causes death via the accumulation of substance Y found in the victim's blood, they will treat the presence of the substance as a diagnostic cue for the presence of a possible cause of the victim's death, poisoning with X. Importantly, however, this diagnostic probability is not the same as the probability that the poison actually caused the victim's death. Generally, an inference of a possible cause based on the observation of its effect, given by the diagnostic probability $P(c|e)$, is not the same as the probability that this cause actually is the singular cause of that effect.[1] Consider, for example, a situation in which C and E are both known to be present. In such a case, a diagnostic query asking for the probability of C's presence is trivial, whereas a singular causation query asking whether c actually caused e is not. One specific example of such a situation is the philosophical preemption scenario about the two rock throwers Billy and Suzy that we described above. Generally, the the diagnostic probability, $P(c|e)$, also contains cases in which C and E co-occurred but the effect was actually caused by an alternative cause (see also Meder et al., 2014). In the special case in which C is a necessary cause of E, both probabilities will be 1.0. We will return to the discussion of diagnostic and singular causation queries in the general discussion.

3.3. *More complex causal mechanisms*

The causal mechanism representations may often be richer than those captured by the causal model in Fig. 2B. For example, reasoners may assume that the causes generate their effects via different possible

[1] We thank an anonymous reviewer for pointing out to us that it is important to distinguish between the two types of inference.

mechanism pathways. As a relevant legal example consider a murder case in which a coroner seeks to determine whether the victim actually died from the bullet that hit them. The general causal relationship between being hit by a bullet and dying can be instantiated in a singular case via different alternative instead of only one possible causal mechanism. In fact, a central goal of an autopsy would be to reconstruct precisely which inner parts of the body have actually been injured by the bullet in the given case, as bullets can lead to death in different ways. Importantly, these different mechanistic possibilities make a singular causal connection between gunshot and victim's death more or less likely. For example, coroners will probably be quite confident that the bullet killed the victim if they find that it went straight to the victim's heart. They will probably be less confident if they discover that the bullet merely damaged some muscle fibers. These conclusions seem warranted because different causal mechanisms imply different causal strengths. Bullets hitting hearts are much more lethal than bullets only damaging muscle fibers.

Knowledge about different causal mechanisms paths can help in the assessment of singular causation not only because different mechanism paths may differ in their causal strengths. Different mechanism pathways also may differ in their in causal latencies, which is also a relevant factor for assessing singular causation relations (cf. Stephan et al., 2018, 2020). For example, bullets going directly to a victim's heart not only have high causal strength, they also manifest their lethal capacity very fast. A short causal latency makes it more likely that a target cause actually caused the target effect because such causes leave less room for alternative causes to preempt them (see also Lagnado & Speekenbrink, 2010). Consider a situation in which coroners not only discover that the victim had been hit by a bullet, but also find traces of poison in the victim's mouth. Was being hit by the bullet or being poisoned the cause of the victim's death? Discovering that the bullet hit the victim's heart should make it more likely that the bullet instead of the poison caused the victim's death because the short causal latency associated with bullets hitting hearts probably left little time for the poison to take effect.

A causal model representing a case in which a reasoner knows that the causes C and A of E generate E via different possible mechanisms is shown in Fig. 2C. Unlike in the previous model shown in Fig. 2B, the causal mechanism variables M_C and M_A this time represent ternary variables that can either be absent (0) or present in one of two different alternative states (1 vs. 2). In this causal model, C and A's causal arrows are assigned two separate causal strength values, w_{cmc1} and w_{cmc2}, representing the strengths of the causes generating values of 1 or 2 in the connected mechanism variables. Similarly, the mechanism nodes M_C and M_A are assigned two different causal strength parameters. For example, the strength parameter $w_{\text{mc1}e}$ denotes the strength with which $M_C = 1$ causes E, and the strength parameter $w_{\text{mc2}e}$ denotes the strength with which $M_C = 2$ causes E.

In our experimental scenario, this causal model was implemented by the two different possibilities of how emergency calls from the castles could be forwarded to the King's palace at the intermediate stations (see Fig. 3). Subjects learned that one possibility are telegraphs ($M_C = 1$ or $M_A = 1$) and that an alternative back-up possibility are pony riders ($M_C = 2$ or $M_A = 2$), which would only be sent out on occasions with insufficient electricity supply for the telegraph towers. We assumed that telegraph towers would intuitively convey the impression of a fast and strong causal process, while pony riders would be considered as a relatively slow causal process. Our scenario description also mentioned that pony riders may fail to reach the palace because they tend to get murdered by bandits living in the forests. This information was supposed to convey the impression of a lower causal strength of the pony riders. Using this information about telegraphs and pony riders, we aimed to establish a causal model in which each of the two possible causes of E operates via one mechanism path that is "superior" on both relevant dimensions, causal strength and causal latency. Our reason for doing so was that we wanted to have a case in which information about one of different possible causal mechanism paths is clearly more useful and relevant than information about another path, analogous to the bullet example where the observation that the bullet struck the victim's heart should lead to much higher confidence that it was the singular cause of death than the observation that the bullet merely damaged some muscle fibers. Our goal here was not to pit causal latency and strength against each other, which was systematically done in Stephan et al. (2020) but to establish clearly "superior" and "inferior" alternative mechanism paths (for an overview on the role of time in causal learning and reasoning, see Buehner, 2017).

To illustrate how our model works for cases with different possible causal mechanisms, we assume that C and A, the two castles in our scenario, are two equivalent competing causes of E whose parameters are matched: $w_{\text{cmc1}} = w_{\text{ama1}}$, $w_{\text{cmc2}} = w_{\text{ama2}}$, $w_{\text{mc1}e} = w_{\text{ma1}e}$, $w_{\text{mc2}e} = w_{\text{ma2}e}$, and so on (see Fig. 2C). Importantly, we assume that the causes' mechanism paths via $M_C = 1$ and $M_A = 1$ are stronger and faster than the alternative mechanism paths leading to E via $M_C = 2$ and via $M_A = 2$.

Fig. 4A shows the model predictions for 28 different singular observations that we tested in Experiment 1. The test cases are listed on the x-axes in the form of neuron diagrams. Green and crossed-out nodes denote active and inactive variables, respectively, and variables whose status is unknown are depicted in black. Green mechanism nodes with a bold contour mean that a telegraph tower is active, the superior mechanism component with higher causal strength and shorter causal latency. Green nodes with a regular contour mean that a pony rider instead of a telegraph was sent. Asterisks mark the target cause c. We chose this large set of test cases because it allows us to systematically test the different components of our model. The set of test cases is divided into five different segments. In all test cases of the first segment the alternative cause A fails to activate its mechanism, while in all cases of the last segment the target cause C fails to activate its mechanism; the cases in the last segment simply mirror those of the first segment. These cases were predicted to elicit the most extreme ratings because we expected that the singular cause of the effect can be determined with certainty in these cases. For all test cases in the first segment in which it was observed that the pigeon sent out from the "competing" castle failed to reach its intermediate station, the competing castle can be ruled out as the singular cause of the alarm in the palace. As we have seen above, our model predicts for such cases that it must have been the target castle that caused the alarm in the King's palace, irrespective of what else is known about the status of the target cause's intermediate mechanism variables. Subjects' judgments should therefore remain uninfluenced by the information about the states of the other variables. For example, in test case 01 it is observed that the Western castle succeeded in activating its superior mechanism component, the telegraph tower, while it merely succeeded in activating its weaker and slower mechanism component, a pony rider, in test case 02. Our model predicts that this difference should be irrelevant since the competing Eastern castle can be ruled out as the singular cause of the observed alarm in the King's palace. As we have seen above, Equation (1) yields values of 1.0 in these cases because the strength parameter of the alternative cause takes on a value of 0 in cases in which it fails to activate its mechanism variables. We included test cases like those in segments 1 and 5 because they allowed us to test if reasoners understand that otherwise crucial differences can be irrelevant under certain conditions.

The third (middle) segment consists of test cases for which the same information is given for the target and the alternative cause, which is why we called them "symmetric" test cases. In these cases, our model predicts maximal uncertainty (i.e., ratings of 0.5) about whether the target cause was the singular cause of the effect. Importantly, this prediction of maximal uncertainty for the symmetric cases is made by our generalized power PC model of singular causation judgments because it incorporates the possibility of causal preemption, captured by $w_c \times w_a \times \alpha$ in Equation (1). For example, on an occasion on which it is known that both castles sent a pigeon, both intermediate stations sent a telegraph, and an alarm occurred in the palace, our model predicts that there is a fifty-fifty chance that the alarm was caused by the target castle (or by the

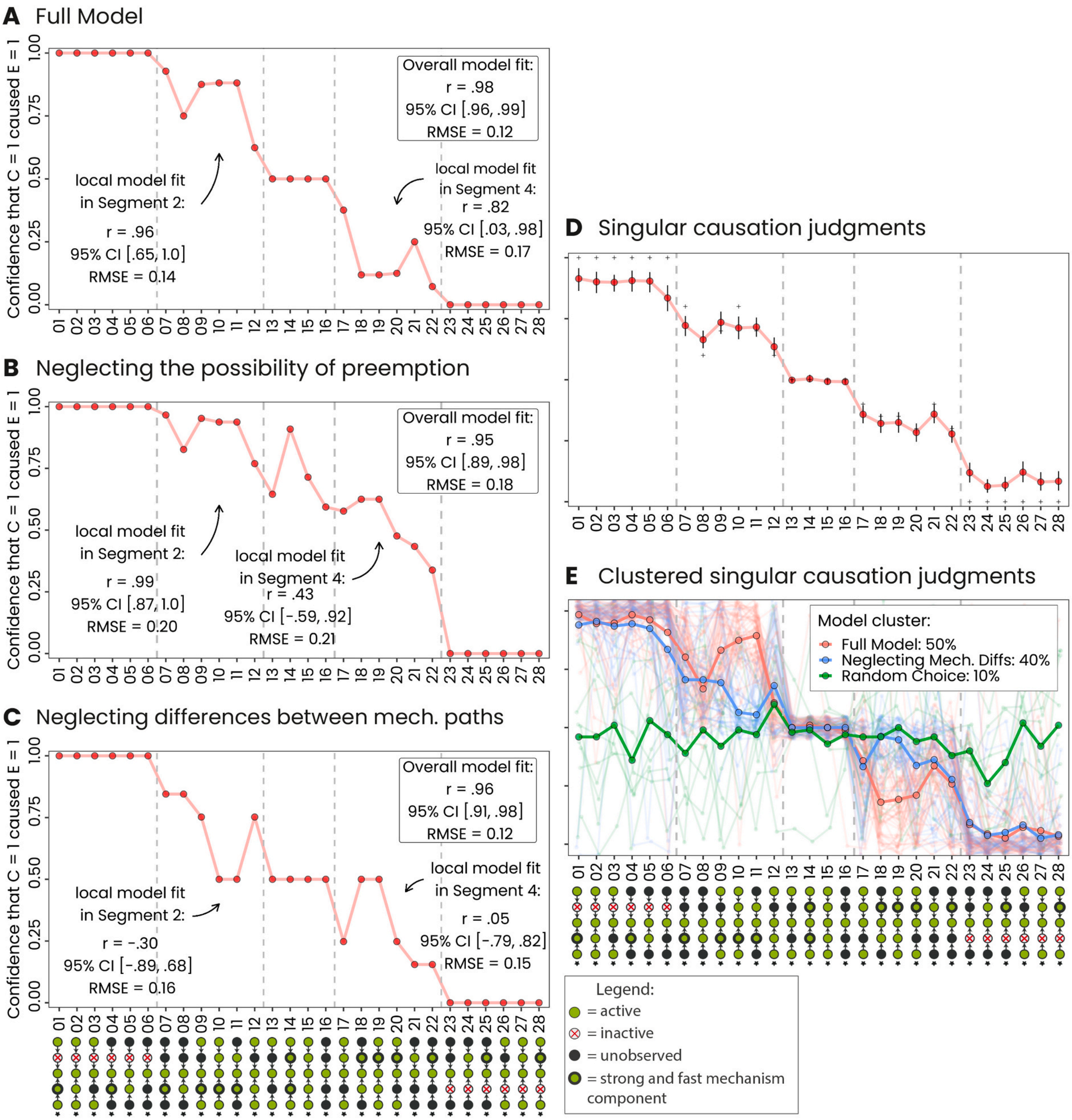

Fig. 4. Singular causation predictions of different models for the 28 singular observations tested in Experiment 1 (A - C), and human singular causation judgments (D - E). *Note.* The "*" in the causal diagrams listed on the x-axes mark the target cause *c*. In panel D, red circles represent means, crosses represent medians, and error bars are 95% CIs. In panel E, bold lines represent means and faint lines represent individual ratings.

competing castle). This prediction rests on the assumption that, in our scenario, the relevant causal parameters, strength and causal latency, are equal for the two candidate causes *C* and *A*. The reason why we included these symmetric test cases is that they best distinguish the predictions of our model from those from its predecessor, the "standard" power PC model of causal attribution proposed by Cheng and Novick (2005) (see also Stephan et al., 2020; Stephan & Waldmann, 2018). The predictions that the standard model would make for our test cases are shown in Fig. 4B. As can be seen there, a consequence of neglecting causal preemption is that that this would lead to the prediction of a target cause preference (or target cause bias). A comparison of the predictions that the two models make for the symmetric test cases of the third segment shows that this target cause preference is particularly pronounced for the symmetric test cases.

The most complicated cases of our test set, which also best demonstrate the relevance of knowledge about different mechanism pathways, are those in segments 2 and 4 (test cases 07 to 12 and test cases 17 to 22) in Fig. 4. These test cases are computationally more demanding because, unlike those in segments 1 and 5 for example, reasoners here need to integrate and aggregate causal strength and latency parameters to make a judgment, and sometimes even need to consider base rates. The specific parameter values we used to compute the predictions were chosen based on our own intuition about the experimental scenarios and the results of a pilot study in which we pre-tested our materials. The data and results of that pilot study can be accessed via our online repository at https://osf.io/325pr/. An R-Script that can be used to reproduce the model predictions for all cases, or to explore how the predictions change for different parameter values is provided at https://osf.io/ud3qb/.

We will now illustrate in more detail how the model works using the test cases of segment 2 (test cases 07 to 12). The test cases in segment 4 (test cases 17 to 22) mirror those of segment 2, and our model here predicts the inverse ratings of segment 2. We assume in our illustration that the target cause marked by the asterisks in the neuron diagrams in Fig. 4 represents the Western castle of our scenario. Our focus will be on the logic behind the model predictions. A supplementary document in which we provide a step by step calculation of the predictions can be found at https://osf.io/wxnvh/.

We first illustrate the predictions for test cases 07 and 08. As can be seen in Fig. 4A, our model predicts that reasoners should be more confident that the Western castle caused the alarm in the King's palace for test case 07 than for test case 08. In both cases, it is unknown whether the two castles sent out a pigeon, and whether the intermediate stations of the competing Eastern castle were active. However, for test case 07 it is observed that the Western castle succeeded in activating its superior mechanism component, the telegraph tower, whereas it merely managed to activate the pony station in test case 08. From the observed activity of the Western castle's intermediate stations in test cases 07 and 08, it can first diagnostically be inferred that the Western castle must have sent a pigeon on these occasions, $P(C = 1|M_C = 1 \wedge M_C = 2) = 1$. Furthermore, since the only cause of activity of the intermediate stations are their respective castles, the probability that the activity of the Western telegraph tower in test case 07 and the Western pony station in test case 08 are caused by the Western castle, $P(c \rightarrow m_C = 1|m_C = 1)$ or $P(c \rightarrow m_C = 2|m_C = 2)$, are also 1.0. This is the case in our scenario because the castles are considered to be necessary causes of their intermediate stations. In a next step, it thus needs to be determined how likely it is that the telegraph (test case 07) or the pony rider (test case 08) caused the alarm. This probability needs to reflect the possibility that the competing castle could have sent a pigeon, and that its intermediate station could have sent a telegraph or a pony rider. The exact probabilities of these events depend on the parameter values of the causal model, that is, the base rate of castle activity, the strength with which the castles activate telegraphs or pony riders, and the probability that one of them is causally preempted by the mechanism variable of the competing cause. The probability that the target Western castle caused the alarm will become smaller, the higher these probabilities are. However, higher singular causation ratings for test case 07 than for test case 08 will be predicted as long as it is assumed that telegraphs (test case 07) have a higher causal strength and shorter causal latency than pony riders (test case 08).

It can be seen in Fig. 4A that the singular causation prediction for test case 09 is a little bit lower than the one for test case 07 but higher than for test case 08. The prediction for test case 09 is lower than for test case 07 because for this test case it is observed that the competing castle actually sent a pigeon. In test case 07, it is possible that the competing castle remained inactive, which also decreases the possibility that it caused the alarm in the palace. For test case 09, by contrast, this possibility does not exist. Here, the competing castle's base rate needs to be neglected, which increases the probability that the competing castle caused the alarm, and simultaneously decreases the probability that the target castle caused the alarm. The magnitude of this decrease between test case 07 and 09 depends on the size of the competing castle's base rate parameter. If the competing castle always sent a pigeon, that is, if its base rate was 1.0, identical predictions would result for the two cases. Finally, the fact that, unlike in test case 07, the target castle is present in test case 09 does not impact on the difference between the predictions for test cases 07 and 09. The reason is that the observed activity of the target castle's telegraph tower in test case 07 implies that the target castle must have sent a message.

For test cases 10 and 11, our model makes identical predictions, although the observed events are not identical.[2] Importantly, the fact that the predictions are the same for these two cases is independent of the specific parameter values that we choose; they are identical because the candidate causes in our scenario are necessary causes of their intermediate mechanism variables. The only difference between test cases 10 and 11 is that the states of the two castles remain unobserved in test case 11. However, since in test case 11 the intermediate stations are active and because these activations can result only if the castles sent a pigeon, the activity of the two castles can be inferred with certainty.[3] Test case 11 again illustrates that the probability of singular causation, $P(c \rightarrow e|e, m_C = 1, m_A = 2)$ and the diagnostic probability of the presence of the cause, $P(c|e, m_C = 1, m_A = 2)$, are not identical. In test case 11, the presence of the target cause can safely be inferred from the presence of its mechanism variable, but it is not certain that the target cause was the singular cause of the effect. To see how the two measures differ for all of our test cases of Experiment 1, a supplementary file in which we pit singular causation predictions and diagnostic probability against each other is provided in our online repository at https://osf.io/3bwyv/.

The prediction with the lowest value in segment 2 is made for test case 12. This is a test case in which both castles sent out a pigeon, the Western intermediate station sent out a pony rider, but it is unknown whether the Eastern intermediate station sent a telegraph or a pony rider, or failed to become active. The reason why this case receives the lowest predicted value of the set is that we set the causal strengths between the castles and telegraph towers (w_{cmc1} in Fig. 2) to a high value. It is thus very likely that the competing Eastern castle successfully activated its superior mechanism component, the telegraph tower. As the target Western castle's intermediate cause would in this case compete with a much stronger and faster intermediate cause (Western pony rider vs. Eastern telegraph), our model predicts that the probability that the Western castle caused the alarm in the palace should be relatively low.

4. Constraints on the utility of mechanism information

So far we have shown why mechanism information is helpful for the assessment of singular causation. However, our model can also be used to identify factors imposing constraints on the usefulness of mechanism information.

4.1. Intersecting causal mechanisms

One such factor is whether the potential causes of the target effect operate via non-intersecting, independent, or via intersecting, dependent mechanisms paths. In the first three causal models shown in Fig. 2 the potential causes operate via non-intersecting paths. The causal model in Fig. 2D represents instead a scenario in which A's mechanism paths intersects with C's. Intersecting causal mechanisms can reduce the utility of mechanism information in the assessment of singular causation. For example, in a situation in which the causal model shown in Fig. 2D represents the mechanisms underlying the causal relations, observing that the target cause's mechanism component was active in a singular case provides no further information about whether c caused e. Since E has no proximate cause other than M_{CMA} in this case, observing e diagnostically implies $M_{CMA} = 1$. Observing $M_{CMA} = 1$ does not give further evidential support for the hypothesis that c caused e.

As a concrete example, consider the version of our experimental scenario that we used in Experiment 2. The two scenarios and their causal structures are illustrated in Fig. 5. The non-intersecting mechanisms condition is shown in Fig. 5A. The scenario description here was very similar to the one of Experiment 1, except that this time we only introduced the telegraph towers as intermediate mechanism components. In the non-intersection mechanisms scenario instantiating the causal model shown in Fig. 2B, the Western and the Eastern castle each transmit their emergency signals via their own intermediate telegraph towers. The intersecting-mechanism scenario instantiating the causal

[2] We thank an anonymous reviewer for pointing out that it is important to elaborate why identical predictions result for these cases.

[3] This is not the case for non-necessary causes, that is, for cases in which a cause's direct child variables have alternative causes. We will discuss such cases further in the section about intersecting causal mechanisms.

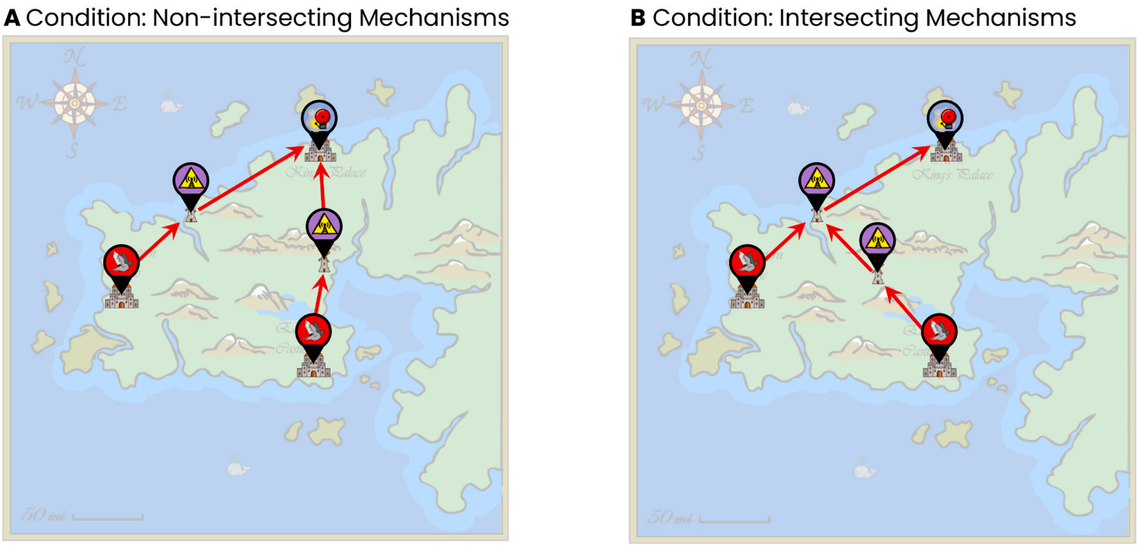
A Condition: Non-intersecting Mechanisms
B Condition: Intersecting Mechanisms
N
W
E
S
King's Palace
50 mi

Fig. 5. Illustrations of the causal structures and their components in the non-intersecting (a) and intersecting mechanisms (b) conditions that were included in the instructions of Experiment 2.

model shown in Fig. 2D is illustrated in Fig. 5B. Here, the intermediate telegraph tower of the Eastern castle, corresponding to node M_A in the causal model, cannot send its telegraphs directly to the palace, corresponding to node E, but has to forward them to the telegraph tower shared with the Western castle, represented by node M_{CMA}. To see how intersecting mechanisms affect the utility of mechanism information, we

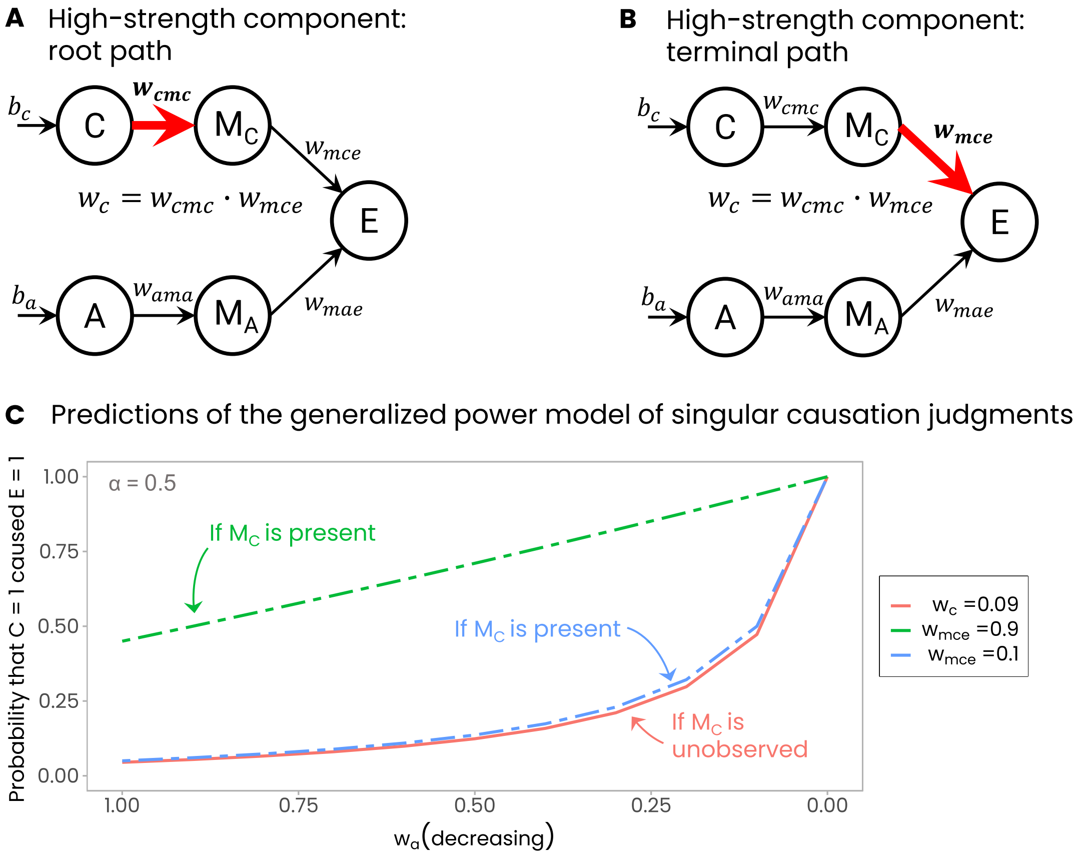

Fig. 6. Illustration showing how the way a cause's global strength (w_c) is distributed across its path components (w_{cmc} and w_{mce}) affects the utility of mechanism information. *Note.* The causal models in A and B represent two cases in which the variables C in the two models have identical global strengths $w_c = w_{cmc} \cdot w_{mce}$, but differ in how w_c is distributed across its components w_{cmc} and w_{mce}. A: The first arrow connecting C and M_C is stronger than the second, that is, $w_{cme} > w_{mce}$. B: The second arrow connecting M_C and E is stronger than the first, that is, $w_{mce} > w_{cmc}$. C: Predictions of the generalized power model of singular causation judgments for the two cases. The red curve shows the predictions for $P(c \to e|c, a, e)$ under the two causal models if M_C is unobserved. In this case, the strength parameter entering into the equation is w_c. The blue and green lines show how the probability that c caused e changes upon observing $M_C = 1$ under the two different causal models. The blue curve represents the causal model shown in A and the green curve represents the causal model shown in B.

assume that the target cause is the Western castle. We now observe that an alarm occurred, that both castles sent a pigeon, but that the status of the telegraph towers is unobserved. Our model predicts that the additional information that the telegraph tower of the Western castle actually sent a telegraph on the given occasion should increase reasoners' confidence that the Western castle caused the alarm in the case of non-intersecting, but not in the case of intersecting mechanisms. In the latter case, observing $M_{CMA} = 1$ is predicted to be irrelevant because its presence is diagnostically implied by the observed presence of the effect. Of course, the causal model shown in Fig. 2D in which A fully exerts its influence via C's mechanism path is an extreme case. We chose this case because it best illustrates the general problem for singular causation judgments arising from intersecting, dependent causal mechanisms.

4.2. Distribution of overall causal strength across causal path components

Another factor constraining the utility of mechanism information that we tested in Experiment 3 is the way in which the strength of the target cause is distributed across its different mechanism path components. Consider again the mechanism model shown in Fig. 2B. An implication of w_c being equal to $w_{\text{cmc}} \cdot w_{\text{mce}}$ is that the probability that c caused e can substantially vary depending on the differences between w_{cmc} and w_{mce} when $M_C = 1$. Fig. 6 shows two causal models that we contrasted in Experiment 3. In both cases C's overall causal strength $w_c = w_{\text{cmc}} \cdot w_{\text{mce}}$ is assumed to be identical, but differently distributed across its two paths components. In Fig. 6A the high-strength component is the root link connecting C to M_C, whereas in Fig. 6B the high-strength component is the terminal path connecting M_C to E.

As an example, consider the version of our fictitious scenario that we used in Experiment 3. The Western and Eastern castles can send out pony riders to the palace, but the pony riders are unable to cover the whole distance. They can only reach the intermediate pony stations positioned halfway. These intermediate stations then send out fresh pony riders who cover the remaining distance. Pony riders not always make it to their destination, however, because they are attacked by evil robber barons. In one version of the scenario that instantiates the causal model shown in Fig. 6A, the pony riders sent out from their castles get attacked only 10% of the time, implying a causal strength of 0.9 for w_{cmc}, whereas those pony riders riding from the intermediate stations to the palace get attacked 90% of the time, implying a strength of 0.1 for w_{mce}. In the other version of the scenario that instantiates the causal model shown in Fig. 6B, these values were simply reversed, $w_{\text{cmc}} = 0.1$ and $w_{\text{mce}} = 0.9$. In both cases, the overall strengths of the castles are identical, for example, $w_c = w_{\text{cmc}} \cdot w_{\text{mce}} = 0.09$.

Despite identical global strength values for w_c under the two causal models, $P(c \to e | c, m_C, a, e)$ turns out to be higher under the causal model in which the strong component of the causal path is the terminal link (B). Fig. 6C shows the predictions of our model for these situations across the range of possible strength values for the competing cause A. First, consider a case in which both castles sent out a pony rider but it is unknown if these riders reached their intermediate stations, i.e., $M_C = ?$ and $M_A = ?$. The probability that an alarm e was caused by the target castle c in this case must be calculated based on C's overall strength parameter w_c. In Fig. 6C, the probability of singular causation for this case is given by the red solid curve. The blue and green curves show how the probability changes upon observing that the target castle's intermediate pony station sent out a rider, $M_C = 1$, in the two different scenarios. As can be seen, observing that the target castle's intermediate station actually sent out a pony rider entails large differences for $P(c \to e | c, m_C, a, e)$ in these two cases. While observing $M_C = 1$ only slightly increases the probability that c caused e if the terminal link is weak (A), the probability increase is much higher if the terminal link is strong (B). Put less formally, our model captures the notion that the stronger the causal link connecting the target cause to its mechanism variable is, the less "surprised" we should be if we actually find this variable to be active in a given situation, and the less our confidence that c caused e should change. If we know that pony riders sent out from the castle often reach their intermediate station, the observation that this has actually happened should have relatively little influence on our confidence that the observed alarm was caused by the target castle.

The example also demonstrates that causal mechanism information becomes less valuable the higher the target cause's overall strength w_c is. The higher w_c is, the less room there is for w_{cmc} and w_{mce} to differ, and the higher the strength of the link connecting C and its mechanism M_C will be. The higher the strength of the root link is, the smaller the impact of observing $M_C = 1$ will be.

We will next present the results of three experiments in which we tested how mechanism information influences people's singular causation judgments. We evaluated to which extent people reason in accordance with our model.

5. Experiment 1

Experiment 1 tested singular causation judgments for the causal structure shown in Fig. 2C, in which the different potential causes can generate the effect via different possible mechanisms. Using this causal model in combination with the 28 different test cases shown in Fig. 4, this experiment provides a comprehensive experimental test of the role of mechanism information in people's singular causation judgments. It allowed us to investigate all relevant components of our model in a single study.

5.1. Methods

5.1.1. Participants

One hundred subjects ($M_{age} = 33.04$, $SD_{age} = 10.74$, 55 male, 43 female, two non-binary) participated in this online experiment and provided valid data. Subjects were recruited from the British online panel *Prolific* (https://www.prolific.co). The inclusion criteria were a minimum age of 18 years, a 90 percent approval rate of subjects' participation in previous studies, "secondary school" as the minimum level of education, and English as native language. To minimize the risk of distraction, subjects were asked to participate only via desktop PC or laptop and not via tablets or smartphones. Prolific workers who had participated in our pilot study were excluded from participation. Subjects were paid £ 2.00 for their participation.

5.1.2. Design, materials, and procedure

The study had a within-subject design in which the 28 test cases were presented one after the other in random order. A demo video of the study can be found at https://osf.io/ycv8u/.

After subjects had read the scenario description about the medieval kingdom Tristonia (see Fig. 3), they had to pass a comprehension test probing their understanding of the relevant aspects of the scenario. Subjects could not proceed to the main task until they answered all instruction check questions correctly.

Each of the 28 test cases was introduced by a short prompt asking subjects to study what had happened in Tristonia on the given day. The singular causation test question was presented on the same screen below the test case. Subjects were asked to indicate how strongly they believed that the alarm in the King's palace which occurred on that day was caused by the Western [Eastern] castle. Responses were given on an eleven-point rating scale whose endpoints were labeled "certain that it was not caused by this castle" and "certain that it was caused by this castle" (the midpoint was labeled "50:50"). Whether the target cause was the Western or the Eastern castle was counterbalanced between subjects. After subjects had finished all observations, they provided demographic information and then finished the study with a short debriefing screen.

5.2. Results and discussion

Subjects' mean singular causation judgments are shown in Fig. 4D. A table with the descriptive statistics can be found at https://osf.io/v9se4/. As can be seen in Fig. 4D, subjects' singular causation judgments overall followed the predictions of our generalized power PC model of singular causation shown in Fig. 4A. Fig. 4A also includes the results of model fits that we computed. Overall, we observed a high fit between model predictions and singular causation judgments. In particular, we found that ratings tended to change when the model predicts they should change, and also that they tended not to change when the model predicts that they should not. For example, for the test cases in the first and last segment, for which our model predicts that subjects should consider different values of the target cause's mechanism variable to be irrelevant, we found that subjects indeed tended to conform to this prediction. An exception here was test case 06 in which the status of the target castle and its mechanism variable both remained unobserved. While subjects gave almost identical mean ratings for the first five test cases, $M_{1:5} = 0.9$, 95% CI[0.86, 0.95], the mean rating for test case 06 ($M_6 = 0.84$, 95% CI[0.79, 0.88]) was slightly lower, $\Delta M = 0.07$, 95% CI[0.03, 0.11]. An analogous pattern was observed in the last segment of test cases. Here, the mirrored test case of test case 06 in segment 1, test case 23, tended to receive slightly higher ratings, $\Delta M = 0.04$, 95% CI[0.001, 0.07]. This pattern suggests that unobserved variable states tended to elicited uncertainty in some participants even in situations in which this should not be the case.

A comparison between the model predictions for the test cases in segments 2 and 4 and the ratings shows that the model overall also accurately predicts subjects' singular causation judgments for these computationally more demanding test cases. For example, subjects' mean judgments tended to follow the predicted u-shaped trend for test cases 07, 08, and 09 in segment 2, and also the predicted inverse u-shaped trend for test cases 20, 21, and 22 in segment 4. Polynomial contrast analyses confirmed that the quadratic trends in segment 1 and 4 were significant. For the quadratic trend in segment 1, the estimate was $D_{quadr.} = 0.13$, 95% CI[0.06, 0.2] ($t(198) = 3.6$, $p < .0005$), and the estimate for the inverse quadratic trend in segment 4 was $D_{quadr.} = -0.16$, 95% CI[-0.22, -0.09] ($t(198) = -4.5$, $p < .0001$). Importantly, these trends document that subjects' singular causation judgments expressed knowledge about different possible mechanism paths via which a cause can lead to its effect. In line with the model predictions for segments 2 and 4, subjects tended to give higher ratings when they observed that the target cause succeeded in activating its mechanism path with higher causal strength and shorter causal latency. However, one characteristic feature of the predicted u-shaped trend that was not observed in subjects' ratings is the predicted difference between test cases 07 and 09 (and, analogously, between test cases 20 and 21). Our model predicts higher ratings for test case 07 than for test case 09. The reason for this prediction is that it is unknown in test case 07 if the competing castle is actually active, whereas it is observed to be active in test case 09. In test case 07, it is less likely that the competing castle caused the effect because it is possible that it was not even active. In test case 07, but not in test case 09, our model incorporates the competing castle's base rate, whereas our subjects seemed to have had the tendency to neglect it.

To further evaluate to which extent subjects' ratings reflected information about different possible mechanisms pathways, we next compared the mean singular causation ratings with those that we would expect if the judgments had not been sensitive to mechanism differences. We therefore computed predictions for a model in which the strengths and the latencies of the mechanisms were identical to compare them with the normative model. Thus, more specifically, we simulated how a reasoner would respond who does not differentiate between pony riders and telegraph towers. The predictions for this scenario are shown in Fig. 4C. As can be seen, this model fails to predict subjects' mean ratings for segments 2 and 4 (though see below), whereas our model predicts the observed trend of the means in these segments well. This analysis thus provides further evidence that, at least on the aggregate level, reasoners tend to be sensitive to causal mechanism information and to differences between different possible mechanism paths.

Another crucial feature of our generalized power PC model of singular causation judgments that distinguishes it from Cheng and Novick's (2005) standard power PC model of causal attribution is that it incorporates the possibility of causal preemption of the target cause by alternative causes of the effect. One implication of this feature is that the generalized model predicts invariant ratings of 0.5 for all symmetric test cases that are listed in segment 3. As Fig. 4D shows, subjects' judgments followed this prediction. The possibility of causal preemption seems to be particularly salient in these symmetric test cases, but preemption of the target cause is also possible in all the test cases of segments 2 and 4. To better assess to which extent subjects' judgments reflect the possibility of preemption in these cases, we compared their judgments not only to our model, but also to the predictions of the standard model, which neglects causal preemption. The predictions of the standard model are shown in Fig. 4C. This model makes similar (but slightly higher) predictions as our preemption-sensitive model for the test cases of segment 2, but its predictions are very different for the mirrored test cases of segment 4. While our model predicts the inverted pattern of segment 2 for these test cases, the standard model neglecting causal preemption does not. Comparing the predictions of the models with subjects' mean singular causation judgments demonstrates that the ratings followed our generalized model. The ratings suggest that subjects took the possibility of causal preemption into account, even in the more complicated asymmetric test cases.

A further finding of the study is that the mean singular causation judgments were overall less extreme than the model predictions. They show the pattern of "weak inferences", a tendency to respond too closely to the midpoint of the scale, which has previously also been documented for other types of causal inferences (see, e.g., Meder et al., 2009; Rottman & Hastie, 2016; see Rottman & Hastie, 2014, for an overview). In a final analysis, we therefore wanted to see whether these weak inferences were obtained because all subjects generally responded too weakly, or because there are different clusters of subjects who differ with respect to the information they used to make their judgments. Subjects individual ratings are shown in Fig. 4E, where it can be seen that they tended to vary a lot. The variability was highest in segments 2 ($SD = 0.21$) and 4 ($SD = 0.19$), which were those for which the predictions of the different models shown in Fig. 4A-C diverged particularly strongly.

To see whether subjects used different reasoning strategies, we conducted a model-based clustering analysis. We included our generalized model (Fig. 4A), the standard model neglecting preemption (Fig. 4B), the predictions of a model neglecting differences between the different mechanism paths (Fig. 4C), and an additional random choice model. The criterion we used to assign subjects to one of the different models was the minimum mean distance of the singular causation judgments from the predictions of the different models. This analysis identified three distinct groups of participants (see Fig. 4E), which indicates that subjects might indeed have relied on different reasoning strategies. The largest cluster (comprising 50% of participants) consisted of subjects whose singular causation judgments were best described by our generalized model (red curve; $r = .99$, RMSE $= .07$). The second largest cluster (comprising 40% of participants) was best described by the cognitively less demanding model that does not take differences between the causes' different mechanism paths into account (blue curve; $r = .99$, RMSE $= .08$). Referring to our experimental scenario, this cluster thus represents subjects who tended not to differentiate between telegraphs and pony riders. One explanation for the existence of this cluster is that some participants may have considered it too difficult to incorporate information about mechanism path differences, and therefore relied on a cognitively less demanding strategy that still incorporates information about the presence or absence of mechanism variables, but not information about different active variable states (and their different strengths and causal latencies). The behavior of ten percent of our participants was best described by the random choice model (green curve), and no subgroup of subjects was identified that systematically neglected the possibility of causal preemption.

A concern that readers might have about this experiment is that we did not have our participants learn the exact parameter values on which the model predictions were based (see Fig. 2C). We thus could have obtained even better, or much worse, model fits if we had used other parameter values.[4] Yet, we do not think that our modeling decision undermines the central conclusions that can be drawn from the study. Our goal was to test if subjects incorporate and integrate information about causal mechanisms, and whether their singular causation judgments are sensitive to differences between different possible mechanisms paths. We think that this study allowed us to test these questions, even if subjects' intuitions about the exact parameter values may have differed to some degree from the ones we used to compute the predictions. Importantly, the predictions for the test cases in the first and last segment, which test whether subjects use mechanism information to rule out potential causes, are insensitive to parameter value changes. Furthermore, the u-shaped and inverted u-shaped patterns of predictions for test cases 07, 08, and 09 in segment 2 and for test cases 20, 21, and 22 in segment 4 are also relatively robust to parameter value changes. We used these test cases to test if subjects were sensitive to differences between possible mechanism paths. Importantly, these cases allow us to test this question as long as the strength parameter values assigned to $M_C = 1$ and $M_A = 1$ are higher than those used for $M_C = 2$ and $M_A = 2$. The telegraph towers and pony riders described in our experimental scenario seemed to have successfully elicited these intuitions, at least in a majority of subjects. Another parameter-size independent characteristic of our model, but not of the standard model neglecting causal preemption, is that it makes inverse predictions for the mirrored test cases. Subjects' singular causation ratings clearly showed this pattern.

In sum, the results of this experiment show that, on average, reasoners incorporate mechanism information in a quite elaborate way, mostly in line with the full generalized power PC model of singular causation judgments. A slight majority of subjects seemed to have understood that knowledge about different mechanism pathways is relevant for the assessment of singular causation because it allows them to use more specifies values of the relevant parameters. Although subjects' mean ratings were overall less extreme than the model predictions, the results show that many subjects considered and systematically integrated a substantial amount of information to derive their judgments. However, we also found that a relatively large number of subjects tended to use a simpler, less demanding reasoning strategy that neglects differences between the different mechanism paths via which a cause can generate its effect. This finding is in line with previous research showing that reasoners often seem to be driven by a need to reduce computational effort (e.g., Fernbach & Rehder, 2013; Waldmann & Hagmayer, 2001). However, even these subjects still considered mechanism information to be crucial for the assessment of singular causation. For example, subjects in this group still used mechanism information in segments 1 and 5 to eliminate one of the potential causes of the target effect.

6. Experiment 2

In Experiment 2 we tested whether reasoners understand that causal mechanism information tends to be less helpful when the potential causes operate via intersecting rather than non-intersecting, independent mechanism pathways. The two causal structures that we tested are the ones that we presented in our theoretical analysis (see Fig. 2B and D).

We compared two test cases in this study (see Fig. 7), one in which the mechanism variables were unobserved (first row of illustrations in Fig. 7) and one in which the target cause's mechanism is observed to be active (second row of illustrations in Fig. 7). The model predictions for the test cases are shown in Fig. 8A. For the non-intersecting-mechanisms condition (blue), the predictions were obtained by setting the strength parameters of the causal structure to $w_{\text{cmc}} = w_{\text{ama}} = 0.6$ and $w_{\text{mce}} = w_{\text{mae}} = 0.9$. While the strength assigned to the telegraph towers (w_{mce} and w_{mae}) was the same as in Experiment 1, we this time assumed higher causal strengths between castles and telegraph towers (w_{cmc} and w_{ama}) because we hypothesized that leaving out the back-up pony riders makes the causal paths leading from the castles to the towers appear more "reliable". Furthermore, since we only mentioned the telegraph towers (i.e., the causal mechanisms of the competing causes that have the same causal latency), the α parameter of our model is set to 0.5, reflecting uncertainty about the causes' preemptive relation whenever both causes are simultaneously sufficient to generate the effect. For the first test case, in which it is observed that both castles sent a pigeon while it is unclear whether their towers sent telegraphs, the model predicts uncertainty about whether the target Western castle caused the alarm. For the second test case, in which the Western telegraph tower is active, the probability that the Western castle caused the alarm is higher because, instead of $w_c = w_{\text{cmc}} \cdot w_{\text{mce}}$, w_{mce} needs to be used as the target cause's strength parameter.

The critical condition is the intersecting-mechanisms condition, in which the observation that the Western telegraph tower is active should be irrelevant. Within the given causal structure, observing the alarm in the King's palace implies that the Western telegraph tower (node M_{CMA} in the causal model shown in Fig. 2D) must have been active, either because it was caused by the Western castle or by the telegraph tower of the Eastern castle. Observing that this telegraph tower is active on an occasion on which an alarm occurred in the palace should therefore not influence singular causation judgments. Subjects' singular causation judgments for both test cases should correspond to the probability that the target instead of the alternative cause activated the causes' shared telegraph tower, represented by the mechanism variable M_{CMA}, which can be calculated using $P(c \rightarrow m_{CMA} | m_{CMA}, c, a)$. The strength parameter values we used were $w_{\text{cmcma}} = w_{\text{ama}} = 0.6$ and $w_{\text{mamcma}} = 0.77$. We used a higher value for w_{mamcma} because we thought that subjects might assume a higher causal strength between two telegraph towers than between a castle and a telegraph tower. α was set to 0.4, which expresses a small preemptive advantage of the target cause. This seems plausible because the alternative cause is only indirectly connected to M_{CMA} via M_A, which implies a longer causal latency. Inserting these values into our model, the probability that target the Western castle caused the observed alarm is $P(c \rightarrow e | e, c, a) = P(c \rightarrow m_{CMA} | m_{CMA}, c, a) = 0.62$.

6.1. Methods

6.1.1. Participants

Eighty-eight subjects ($M_{age} = 31.39$, $SD_{age} = 11.47$, 44 male, 43 female, one non-binary), recruited via *Prolific*, participated in this online experiment and provided valid data. The inclusion and exclusion criteria were the same as in the previous study. Subjects received £ 1.25 for their participation. The rationale behind the sample size was that it allows us to reliably detect the interaction effect resulting from the model predictions. The interaction effect shown in Fig. 8A corresponds to a difference of differences of about $\Delta D = 0.2$. With a sample size of $N = 88$, the 95% CI for $\Delta D = 0.2$ would on average be [0.09, 0.31], assuming that the *SD* of the singular causation ratings is 0.18 (which is the mean *SD* measured in Experiment 1).

6.1.2. Design, materials, and procedure

The study had a 2 (dependency of mechanisms: non-intersecting vs. intersecting mechanisms; between subjects) × 2 (observed singular case: $[C = 1, M_C = ?, E = 1, M_A = ?, A = 1]$ vs. $[C = 1, M_C = 1, E = 1, M_A = ?, A = 1]$; within subject) mixed design. A demo video can be found at htt

[4] We thank Sam Johnson for pointing this problem out.

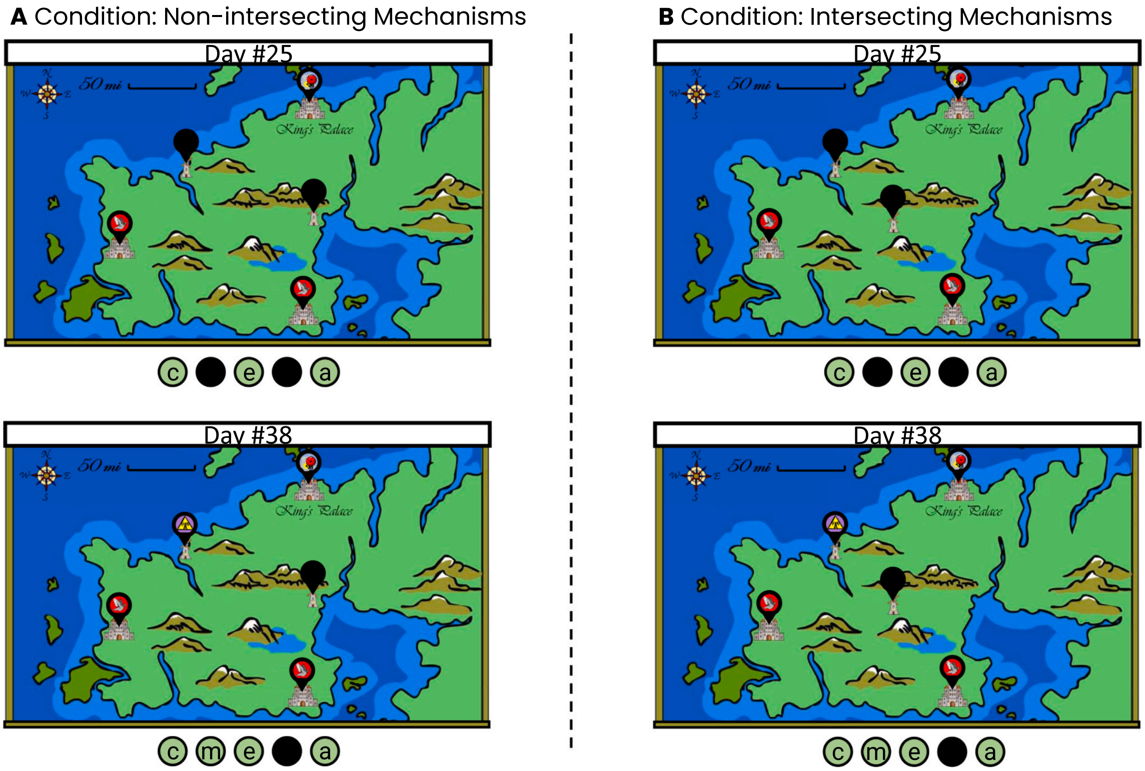
A Condition: Non-intersecting Mechanisms
B Condition: Intersecting Mechanisms
Day #25
Day #38
50 mi
King's Palace
c
m
e
a

Fig. 7. The two test cases shown to participants in the non-intersecting (A) vs. the intersecting mechanisms (B) condition of Experiment 2. *Note.* The diagrams below each test picture illustrate the instantiated variable values of the underlying general causal structure: Green nodes = active, and black nodes = unobserved. Subjects only saw the test pictures, but not these diagrams.

ps://osf.io/w2rph/.

As in Experiment 1, subjects had to pass a comprehension test prior to the test phase. For this experiment it was particularly important that subjects understood the instructed causal structure precisely, especially that the effect cannot be generated by any other than the instructed causes. One comprehension check question therefore asked subjects to indicate the probability of an alarm in the palace if none of the castles sent an emergency call, and all the telegraph towers remained inactive. Subjects could only proceed to the test phase if they answered zero to this question.

The presentation order of the two test cases was counterbalanced between subjects. Subjects indicated how strongly they believed that the alarm in the King's palace was caused by the Western castle. To prompt subjects to reason thoroughly about the cases, we this time also asked them to provide brief justifications for their ratings.

Next we asked subjects an additional diagnostic probability query in the end, which referred to the castles' shared telegraph tower, M_{CMA}. We again showed subjects the test case in which the state of the telegraph tower was unobserved ($M_{CMA} = ?$), and asked them to indicate the probability that it is active in this case ($M_{CMA} = 1.0$), $P(m_{CMA}|e)$. This question allowed us to see if subjects in the intersecting-mechanisms condition correctly inferred that observing the alarm implies that the telegraph tower is active.

6.2. Results and discussion

Fig. 8B shows the mean singular causation judgments. As can be seen there, we observed the expected interaction pattern. As predicted, when subjects in the non-intersecting-mechanisms condition saw the test case in which the status of the Western telegraph tower was unobserved ($M_C = ?$), they were uncertain whether the Western castle had caused the alarm. They gave higher ratings, however, when they saw the test case in which the Western telegraph tower had sent a signal ($M_C = 1$), $M_{Diff} = 0.36$, 95% CI [0.27, 0.44]. The results in this condition replicate the finding of Experiment 1 that observed mechanism variables lead to increased confidence that a target cause was the singular cause of an observed effect.

Subjects in the critical intersecting-mechanism condition gave higher ratings than subjects in the non-intersecting-mechanisms condition when the status of the Western telegraph tower was unobserved ($M_{CMA} = ?$). Furthermore, in line with the model predictions, they gave lower ratings than subjects in the independent mechanisms condition for the test case in which the Western telegraph tower had sent a signal ($M_{CMA} = 1$).

Subjects thus demonstrated that they understood that the observed presence of a mechanism variable connecting a cause to its effect provides more support for a singular causal connection between target cause and effect when this mechanism operates independently of potential alternative causes of the target effect. The interaction effect corresponded to the prediction, $\Delta D = 0.19$, 95% [0.07, 0.31]. However, in contrast to the normative predictions, subjects in the dependent mechanism condition still increased their singular causation ratings upon observing $M_{CMA} = 1$, $M_{Diff} = 0.17$, 95% CI [0.08, 0.25]. Thus, their understanding of this situation was imperfect.

One explanation for the observed confidence increase in the intersecting-mechanisms condition could be that some subjects did not understand that the Western telegraph tower must have caused the alarm on this occasion: $P(m_{CMA}|e) = 1.0$ in this case. To check this possibility, we analyzed subjects' ratings for the additional diagnostic probability query in which we asked them to indicate how likely it is that the Western telegraph tower is active. In line with what the causal structures imply, subjects in the intersecting-mechanisms condition gave higher ratings than subjects in the non-intersecting-mechanisms condition, $M_{Diff} = 0.73 - 0.49 = 0.24$, 95% CI [0.14, 0.34], but their ratings were too low. We also found that the mean diagnostic probability rating of 0.73 resulted from two distinct group of subjects: Half of the

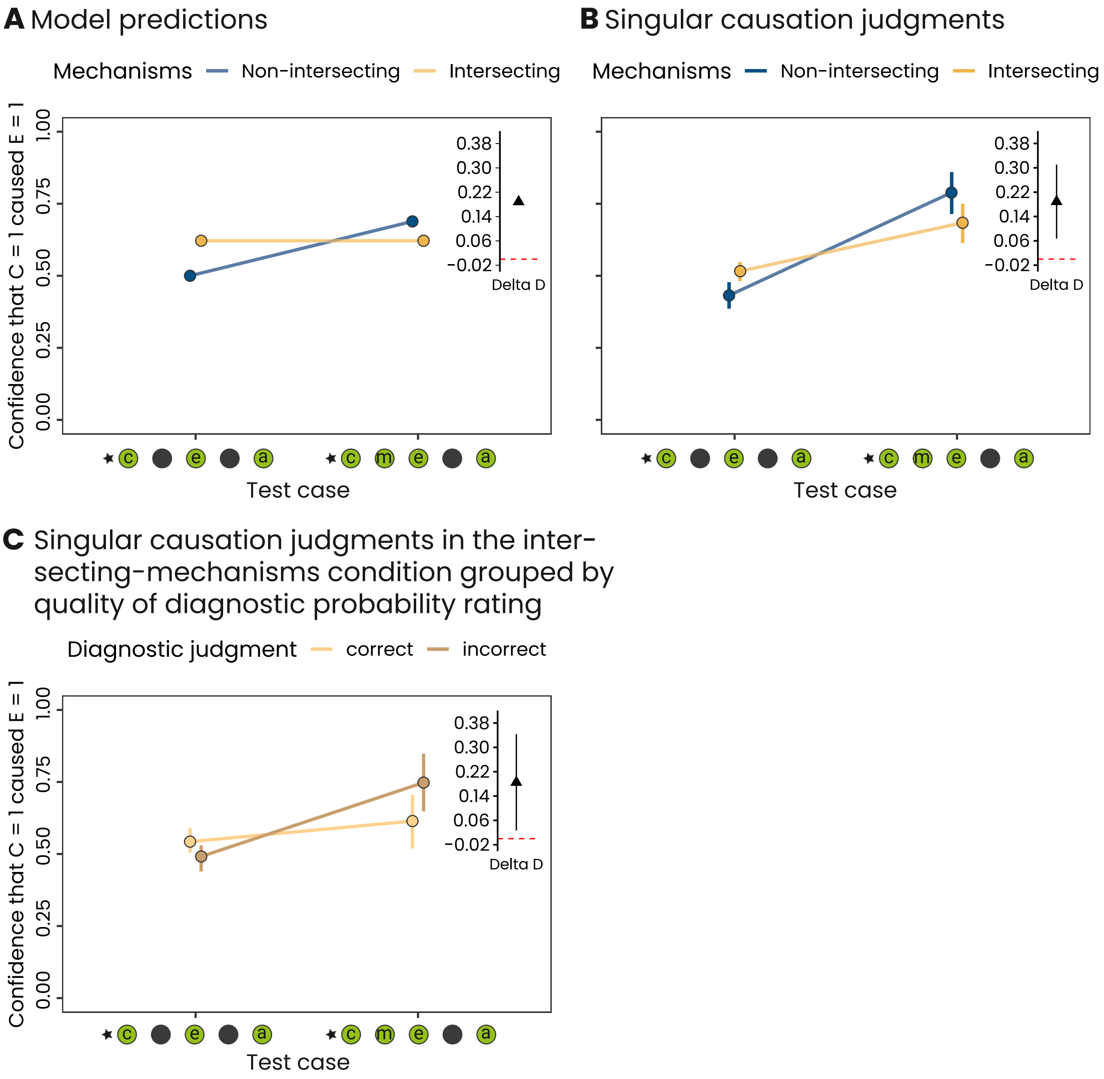

Fig. 8. Model predictions for and results of Experiment 2. *Note.* The results (panels B and C) show mean singular causation judgments. Error bars denote 95% CIs. Difference plots in the right area of each graph show the interaction effect. The diagrams on the x-axes show the test cases in the form of neuron diagrams (cf. Fig. 7). Green nodes = active variable, and black nodes = unobserved variable. Asterisks mark the target cause *c*.

subjects gave correct ratings of 1.0, but the other half gave lower ratings and thus seemed to not have understood that $E = 1$ implies $M_{CMA} = 1$. The majority of these participants reported to be uncertain (i.e., they gave ratings of 0.5).

To see whether these subjects were responsible for the observed positive slope in the singular causation judgments in the intersecting-mechanisms condition (yellow line in Fig. 8B), we compared subjects who gave $P(m_{CMA}|e)$ ratings of 1.0 with those who provided values < 1. Fig. 8C shows that the difference in the singular causation ratings in the intersecting-mechanisms condition was indeed largely driven by subjects who failed to realize that $P(m_{CMA}|e) = 1$ (dark line). These subjects increased their confidence that *c* caused *e* upon observing $M_{CMA} = 1$, $M_{Diff} = 0.75 - 0.49 = 0.26$, 95% CI [0.14, 0.37]. There was no significant increase for subjects who made correct diagnostic judgments, $M_{Diff} = 0.61 - 0.54 = 0.07$, 95% CI [-0.04, 0.18]. The difference of differences was $\Delta D = 0.19$, 95% CI [0.03, 0.34].

The results of this experiment show that reasoners are on average aware that the degree to which causal mechanism information supports singular causation judgments depends on whether the causal mechanisms of the potential causes are statistically independent or not. Many subjects understood that observing independent mechanism variables is more useful than observing dependent mechanism variables that can be caused by alternative causes. Not all participants comprehended the different structures fully, though. Our results suggest that some reasoners seem to have difficulties realizing that observing $E = 1$ is of higher diagnostic value for the presence of the target mechanism if the causal mechanisms intersect, and that the explicit observation of the target mechanism therefore provides less support for the hypothesis of a singular causal connection between target cause and effect.

7. Experiment 3

The goal of Experiment 3 was to test if reasoners understand that the utility of observing a cause's mechanism variable also depends on how the cause's overall strength is distributed across its different mechanism components. The two causal models and the parameter values that we tested in this study are the ones that we presented in our theoretical analysis (see Fig. 6). If reasoners understand how the distribution of a target cause's overall causal strength w_c across its path components (w_{cmc} and w_{mce}) affects the utility of mechanism information, subjects who learn that w_{mce} is the strong component should increase their confidence that *c* caused *e* more upon observing $M_C = 1$ than subjects who learn that w_{mce} is the weak component.

7.1. Scenario and predictions

The experimental scenario in this experiment only introduced pony riders as transmitters of emergency signals. It was pointed out that pony riders do not always reach their destinations because of occasional attacks by evil robber barons living in Tristonia's forests and mountains. To convey the impression that path sections differ in causal strength, we introduced "high-danger zones" in which attacks by evil robber barons were particularly likely. The illustrations we used are depicted in Fig. 9. In the scenario instruction presented in the high-strength component = terminal path condition (Fig. 9A), it was mentioned that 90 percent of pony riders sent out by their castles are attacked and murdered (implying a causal strength of $w_{\mathrm{cmc}} = 0.10$). Pony riders sent out from the intermediate stations were described as having only a ten percent risk of being attacked and murdered (i.e., $w_{\mathrm{mce}} = 0.90$). In the high-strength component = root path condition (Fig. 9B) the instructed risk values were reversed.

As in Experiment 2, subjects were presented with two test cases, which are represented by neuron diagrams on the graphs' x-axes in Fig. 10A-B. In one test case, subjects were informed that both castles had sent out a pony rider, but whether or not these pony riders arrived at their intermediate stations was unknown. In the other test case, subjects additionally learned that the target castle's intermediate pony station had sent out a pony rider. The model predictions are shown in Fig. 10A. They are based on the instructed causal strengths. The value of our model's α parameter was set to 0.5. Since the overall causal strengths of the causes are identical under the two instructed causal models ($w_c = w_a = 0.9{\cdot}0.1 = 0.09$), $P(c \rightarrow e|c, a, e)$ equals 0.5 for the first test case in which $M_C = ?$ and $M_A = ?$ Upon observing $M_C = 1$, the probability that c caused e should increase more in the high-strength component = terminal path than in the high-strength component = root path condition. In the former condition, the target causal strength inserted into the model becomes $w_{\mathrm{mc}} = 0.9$, while in the latter it becomes $w_{\mathrm{mc}} = 0.1$. Thus, $P(c \rightarrow e|c, m_C, a, e) = 0.95$ and $P(c \rightarrow e|c, m_C, a, e) = 0.53$ in the high-strength component = terminal path and the high-strength component = root path conditions, respectively.

7.2. Methods

7.2.1. Participants

One hundred and sixty subjects ($M_{age} = 33.02$, $SD_{age} = 12.11$, 61 male, 82 female, one person indicated "other") recruited via *Prolific* participated in this online experiment and provided valid data. Subjects received £ 1.25 for their participation. The rationale behind the sample size was that we intended to be able to reliably detect relatively small effects. Although our model predicts an interaction effect of *Delta D* = 0.47, we conservatively based our sample size calculation on a smaller interaction effect of *Delta D* = 0.1. With a sample size of $N = 160$ the 95% CI for $\Delta D = 0.1$ can be expected to be [0.02, 0.18] (assuming again $SD = 0.18$ for the singular causation ratings).

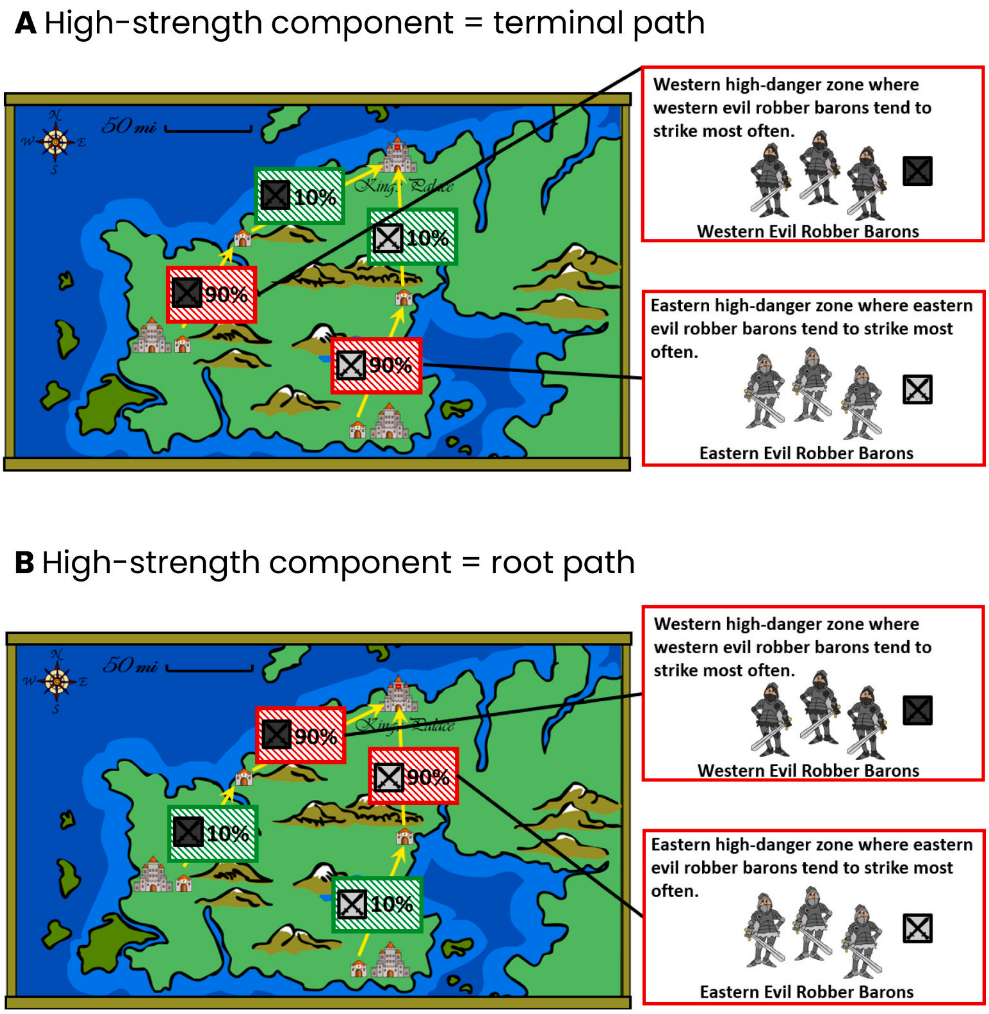
A High-strength component = terminal path
N
W
E
S
50 mi
King's Palace
10%
10%
90%
90%
Western high-danger zone where western evil robber barons tend to strike most often.
Western Evil Robber Barons
Eastern high-danger zone where eastern evil robber barons tend to strike most often.
Eastern Evil Robber Barons
B High-strength component = root path
N
W
E
S
50 mi
King's Palace
90%
90%
10%
10%
Western high-danger zone where western evil robber barons tend to strike most often.
Western Evil Robber Barons
Eastern high-danger zone where eastern evil robber barons tend to strike most often.
Eastern Evil Robber Barons

Fig. 9. Scenario illustrations shown in Experiment 3. *Note.* A: Illustration shown in the 'high-strength components = terminal paths condition'. B: Illustration shown in the 'high-strength components = root paths condition'.

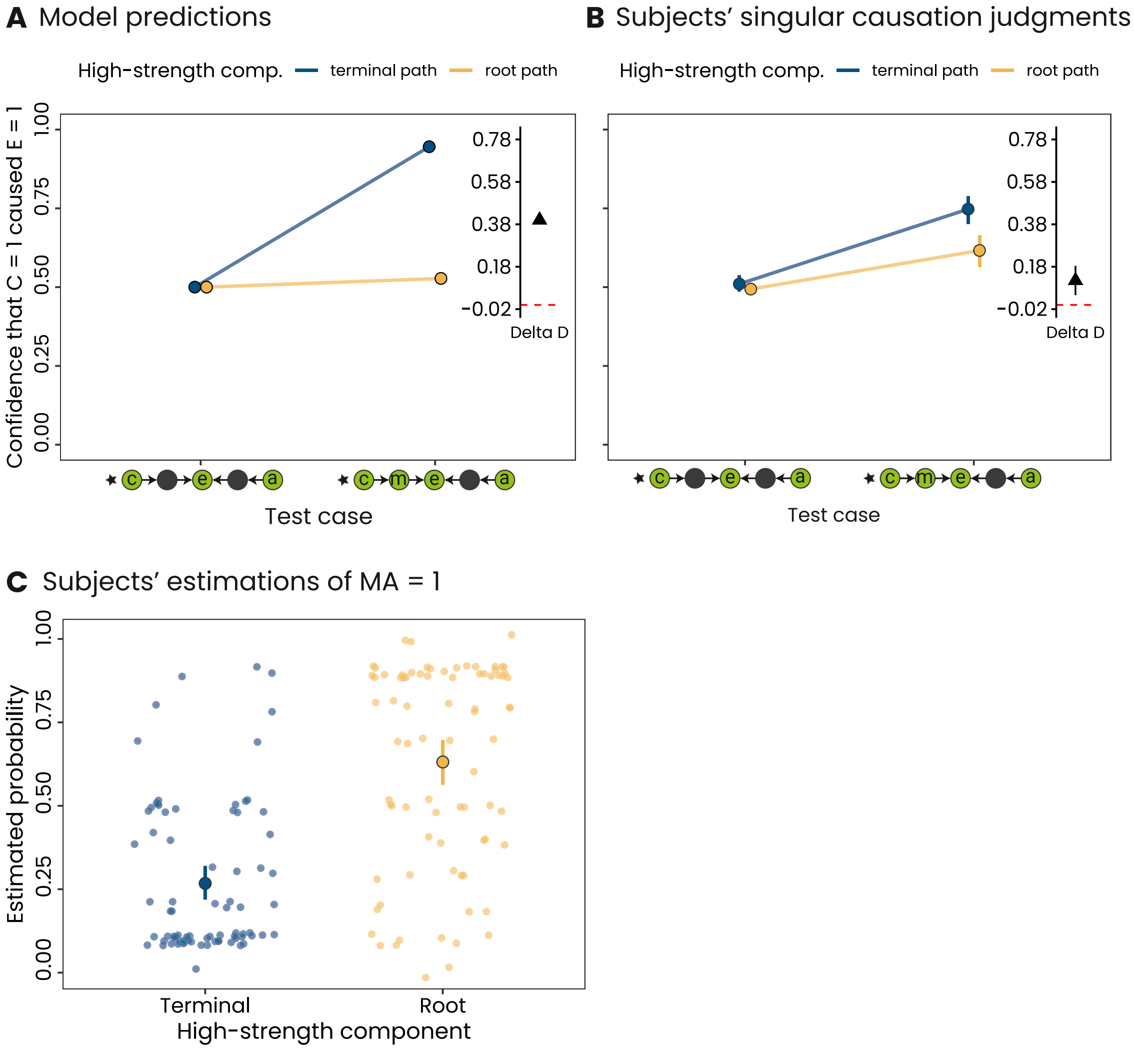

Fig. 10. Model predictions for and results of Experiment 3. *Note.* Error bars denote 95% CIs. The difference plots in A and B show the interaction effect. C: Ratings for the additional test question asking for the probability of the presence of the alternative cause's mechanism variable M_A. The diagrams on the x-axes show the test cases in the form of neuron diagrams. Green nodes = active variable, and black nodes = unobserved variable. Asterisks mark the target cause *c*.

7.2.2. Design, materials, and procedure

The study had a 2 (high-strength component of causal paths: root paths vs. terminal paths; between subjects) × 2 (observed singular case: $[C = 1, M_C = ?, E = 1, M_A = ?, A = 1]$ vs. $[C = 1, M_C = 1, E = 1, M_A = ?, A = 1]$; within-subject) mixed design. A demo video of the procedure can be found at https://osf.io/fum5j/.

The order in which the test cases were presented was counterbalanced between subjects. After subjects had given their singular causation judgments, we asked them an additional question that allowed us to see if our causal strength manipulation was successful. We showed subjects again the test case in which the status of the competing castle's intermediate pony station is unobserved (the second case shown on the x-axes of Fig. 10A-B). We asked subjects to estimate the probability that the unobserved pony station sent out a pony rider in this case. If subjects correctly learned the instructed strength values, subjects in the high-strength component = root path condition should rate $P(M_A = 1) = 0.9$, whereas subjects in the high-strength component = terminal path condition should rate $P(M_A = 1) = 0.1$.

7.3. Results and discussion

Fig. 10B summarizes the results. For the first test case, participants in both conditions indicated to be uncertain ($M = 0.494$, 95% CI [0.458, 0.530] vs. $M = 0.510$, 95% CI [0.474, 0.546] in the high-strength components = root paths and the high-strength components = terminal paths conditions, respectively). Upon observing $M_C = 1$ subjects in the high-strength components = terminal paths condition increased their singular causation judgments more than subjects in the high-strength components = root paths condition ($M = 0.748$, 95% CI [0.712, 0.783] vs. $M = 0.616$, 95% CI [0.580, 0.652]). The interaction effect was $\Delta D = 0.115$, 95% CI [0.046, 0.184], which is smaller than predicted by the model. As in Experiments 1 and 2, subjects showed weak inferences, despite the fact that we explicitly instructed the relevant causal strength parameters.

Ratings for the additional question asking for the probability of $M_A = 1$ are shown in Fig. 10C. Subjects in the high-strength components = root paths condition gave higher probability ratings ($M = 0.631$, 95% CI [0.563, 0.699]) than subjects in the high-strength components = terminal paths condition ($M = 0.268$, 95% CI [0.216, 0.319]), $M_{diff} = 0.364$, 95% CI [0.279, 0.448]), indicating an overall effective experimental manipulation: subjects who learned that the first causal links between the castles and their intermediate stations have high (i.e., 0.9) causal strengths indicated much higher probabilities for $M_A = 1$ than subjects who learned that these initial links were weak. However, the differences between the conditions were smaller than expected based on the instructed parameter values. This, in turn, explains why the observed interaction effect for the singular causation ratings was weak. Moreover, we also observed large interindividual differences in these ratings, as shown by the jittered dots in Fig. 10C.

To obtain further evidence that the observed interaction effect for subjects' singular causation ratings is driven by subjects who reasoned in accordance with our model, we analyzed the correlation between subjects' $P(M_A = 1)$ ratings and their singular causation ratings. The rationale for this analysis was that subjects who think that the activation of the intermediate pony stations is unlikely should increase their singular causation judgments more when observing that the target castle's intermediate station had sent out a pony rider than subjects who think that activation of the intermediate pony stations is likely. We thus should see an overall negative correlation between $P(M_A = 1)$ and $P(c \rightarrow e|c, m_C, a)$ ratings, which is what we found, $r = -0.174$, 95% CI [$-$ 0.320, $-$ 0.0190].

In sum, the results of this experiment suggest that reasoners on average tend to understand that the way in which the causal strength between a target cause and its effect is distributed among its different causal-path components constrains the utility of mechanism information for singular causation judgments. Subjects tend to understand that if a cause with an overall strength of $w_c = w_{\text{cmc}} \cdot w_{\text{mce}}$ is likely to activate its mechanism component (high value of w_{cmc}), observing that its mechanism component is indeed active is less relevant for the singular causation judgment than if a cause only rarely activates its mechanism component (low value of w_{cmc}). This study thus demonstrates an interesting case in which reasoners, despite identical representations of a target cause's overall causal strength (w_c), come to very different conclusions when estimating the probability of a singular causal link between target cause and effect. It not only matters what is known about the overall strength of a target cause, but also how the overall strength is allocated to the different sub-links of the mechanism path connecting the cause with its effect.

8. General discussion

Causal mechanism information has been regarded as an important cue for establishing singular causation relations. The philosopher of science Nancy Cartwright (2017), for example, lists the discovery of intermediate steps in a causal sequence as one of the crucial indicators of singular causation. Several previous studies (e.g., Ahn et al., 1995; Johnson & Keil, 2018) showed that lay people search for causal mechanism information when they are asked to determine the cause of a particular event. What has been missing so far is a formal computational account explaining why causal mechanism knowledge is relevant for singular causation judgments. Based on the power PC framework of singular causation judgments (Cheng & Novick, 2005; Stephan et al., 2020; Stephan & Waldmann, 2018), we here presented such a formal model. Moreover, we systematically assessed to which extent reasoners are sensitive to the different factors identified by our model.

Experiment 1 demonstrated that singular causation judgments are overall in line with the predictions of the model. A very robust normative pattern that we found was that reasoners use causal mechanism information to engage in eliminative reasoning (Bird, 2005): subjects used information about the inactivity of mechanism components to rule out potential causes of an effect. Experiment 1 also documented that reasoners are sensitive to the more subtle predictions of our model. The observed singular causation judgments indicate, for instance, that reasoners understand the roles of causal strength and temporal information in causal mechanisms. However, while many of our participants were sensitive to the different factors included in our model, we also found that a number of subjects seemed to have relied on simplified strategies neglecting crucial components of the model.

We also tested subjects' sensitivity to factors constraining the epistemic utility of causal mechanism information. Experiments 2 and 3 showed that reasoners tended to understand the relevant constraints. Most of our subjects in Experiment 2 understood that observing a target cause's mechanism variable is less informative in the context of intersecting mechanism paths, in which the target cause's mechanism variable can be activated by the alternative cause, and in which the observed presence of the effect diagnostically implies the presence of the target mechanism. Experiment 3 showed that subjects also tend to understand that the degree to which an observation of a target cause's mechanism variable supports the conclusion of a singular causal connection between target cause and effect depends on how the target cause's overall causal strength is distributed across its mechanism path components. Yet, even though we tested prototypical scenarios in which mechanism information should clearly be uninformative, in both experiments we also found that a number of subjects continued to incorporate mechanism information in their singular causation judgments in these cases. This finding is consistent with earlier research which showed that reasoners often process tasks superficially (see, e.g., Stephan et al., 2021; Waldmann, 2000, 2001). This tendency may be even stronger if the relevant information is conveyed via descriptions, as was the case in our studies, rather than by experience (cf. Rehder & Waldmann, 2017).

A key finding of our study is that mechanism information is useful in the assessment of singular causation because it allows reasoners to insert more specific values for two crucial parameters, causal strength and causal latency. Our findings complement previous studies that documented the important role these parameters play for other types of causal inference and for causal learning. For example, previous studies have shown that causal strength knowledge is used in predictive, diagnostic, and interventional judgments (see, e.g., Fernbach et al., 2011; Meder & Mayrhofer, 2017a; Meder et al., 2014; Rottman & Hastie, 2014, for overviews). Similarly, several previous studies documented that causal latency intuitions play an important role in causal cognition. For example, intuitions about the time it takes a cause to produce an effect not only shape people's singular causation judgments, but also play a central role in how people induce general causal relationships from observed patterns of covariation (for an overview, see Buehner, 2017).

8.1. Limitations and directions for future research

To the best of our knowledge, our study represents the most systematic and comprehensive test of the role of mechanism information in singular causation judgments. However, there are also limitations that need to be addressed in future studies. The probably most obvious limitation is that we have analyzed and tested only situations in which a target cause C competes with a single known alternative cause A of the effect. In the vast majority of real-life scenarios, reasoners must deal with the fact that there might exist a multitude of further unknown causes of the effect. Our generalized power PC model of singular causation judgments can be applied to such contexts as well (see Stephan et al., 2020, where this is discussed in the General Discussion). The model is not restricted to cases in which A represents a single known alternative cause. Instead of using separate values for the base rate b_A and the causal strength w_A of the alternative causes, we can insert the probability of the effect in the absence of the target cause, $P(e| \neg c)$ in such cases. This probability reflects the joint influence of all (known and unknown) alternative causes of the effect (see Cheng, 1997; Griffiths & Tenenbaum, 2005). An interesting question, however, is what the unknown background causes imply for the utility of mechanism information in singular causation judgments. One obvious problem that arises in such contexts is that we cannot rely on information about the alternative causes' mechanism variables anymore.[5] If we do not know what the alternative causes are, we also cannot know via which mechanism variables they generate the effect. This, in turn, implies that we cannot use information about the absence of mechanism variables to rule out the unknown alternative causes as singular causes of the effect. Contexts

[5] We thank Sam Johnson for the suggestion to discuss the implications of such cases.

with unobserved alternative causes imply that we can never be certain that c actually caused e. Our model captures this fact because $P(c \rightarrow e|c, e)$ will be < 1.0 if $P(e| \neg c) > 0$.[6]

Information about mechanism variables that belong to the target cause can still be relevant in such cases, however. We have shown that $P(c \rightarrow e|c, m_C, e)$ tends to be higher than $P(c \rightarrow e|c, e)$ because the causal strength parameter for M_C, w_{cmce}, that needs to be used in this case will on average be higher than C's global strength parameter w_c that we must use if M_C is unobserved. Whether or not there exist multiple unknown alternative causes of the effect is irrelevant in this case. Even with unknown background causes reasoners should continue to search for mechanism information.

A more severe problem in contexts with unknown background causes is the calculation of the α parameter of our model, which is relevant for computing the probability of causal preemption (see also Stephan et al., 2020). For unknown background causes, we cannot know if and when they occurred in the target situation; their onset times as well as their causal latencies will be unknown to us. We thus cannot know how likely these unknown causes preempt the target cause. One possible solution in such contexts is to set α to a conservative value. One option would be to set $\alpha = 1.0$, which would reflect the assumption that alternative causes always preempt the target when they are sufficiently strong to generate the effect (as estimated by $P(e| \neg c)$). $P(c \rightarrow e|c, m_C, e)$ would then be a conservative estimate because it would represent the lower boundary of the probability that c is the singular cause of e. Future studies need to study how reasoners make causal inferences in these situations.

On a conceptual level, a criticism of our study might be that we relied on a very specific characterization of causal mechanisms. We here adopted the causal Bayes net view on causal mechanisms (Pearl, 2000), according to which causal mechanisms reduce to dependencies between variables in a more fine-grained network relating a cause to its effect. There exist other views though (see Johnson & Ahn, 2017, for an overview). For example, according to Stuart Glennan, a proponent of the "new mechanical philosophy", causal mechanisms consist of "entities (or parts) whose activities and interactions are organized so as to be responsible for the phenomenon" (Glennan, 2017, p. 17). According to this view, mechanism are more than mere statistical dependencies between variables. Other theories based on the theory of force dynamics assume that mechanism involve the notion of force transmission between physical objects who play the roles of agents and patients (e.g., White, 1989; Wolff, 2007). Distinguishing between these different theories is beyond the scope of this paper, but we believe that alternative accounts of mechanisms would make similar predictions (see also Stephan et al., 2021). We chose the causal Bayes net approach because, unlike many of its alternatives, it combines structural modeling with quantitative parameterizations, which allows us to derive precise numeric predictions from the models.

If we look again at the singular causation judgments observed in Experiment 1, we see that they were altogether slightly less extreme than the model predictions. One possible explanation we have given for this pattern is that different subjects may have assumed different parameter values. An additional explanation for this tendency to provide conservative ratings, which was corroborated by our model-based cluster analysis, is that some subjects neglected mechanism differences, which in our cases implies lower values for the test cases in segments 2 and 4 of our test set. A further factor might be that subjects' judgments also may have reflected a certain degree of uncertainty about the exact parameter values. The predictions of our model did not incorporate parameter uncertainty, which would require the model to use parameter distributions instead of point estimates (cf., Lu et al., 2008; Meder & Mayrhofer, 2017a; Meder et al., 2014; Stephan & Waldmann, 2018). An interesting next step would be to see if the incorporation of parameter uncertainty would increase the model's predictive accuracy.

In all our experiments subjects were asked to evaluate different test cases that were presented to them. Subjects could not control which information about the target situation they would actually like to receive. An interesting avenue for future studies would therefore be to use more active tasks in which subjects can decide themselves whether or not they would like to consult mechanism information. Such tasks would be particularly interesting for scenarios like the ones we tested in Experiments 2 and 3, for which, according to our model, the utility of mechanism information is assumed to be constrained under certain conditions. In everyday life, the search for additional information typically costs effort and time. Thus, it would be interesting to see if more effortful and time consuming active learning tasks would prompt subjects to think more thoroughly about these scenarios.

In our studies, we only used one specific experimental scenario. We thought that a scenario about a medieval kingdom would be particularly engaging and keep participants motivated. However, it must be kept in mind that our scenario about information transmission via carrier pigeons and telegraph towers might have triggered a rather mechanical construal of causality. As has been shown by Lombrozo (2010), lay people's causal ascriptions may differ depending on whether the described events are construed mechanistically with a focus on the underlying mechanical and physical processes or teleologically with a focus on the goals and functions of the system. Moreover, Johnson and Keil (2018) have found that people's inclination to look for mechanism information seems to be particularly pronounced in physical causation, while it seems to be less strong in social causation. Our model is domain general and would make similar predictions in these different scenario types if the underlying causal models are identical. A possible explanation of differences is therefore that people may systematically use different causal models for different domains. Future studies should test a broader range of scenario types to test the robustness of our model and to increase the external validity of our findings.

The present research focused on singular causation queries, which ask whether an observed effect event was actually caused by a target cause. A related type of query that has been investigated in several previous studies are diagnostic probability queries (Fernbach et al., 2011; Meder & Mayrhofer, 2017a; Meder et al., 2014; see also Meder & Mayrhofer, 2017b; Waldmann et al., 2008b). Diagnostic judgments start with an observation of an effect. However, as we have discussed in the theory section of this article, unlike singular causation queries diagnostic queries merely ask for the probability of the presence of a target cause, $P(c|e)$, and not whether an observed present target cause was actually causally responsible for the effect, $P(c \rightarrow e|c, e)$. As we have discussed, causal mechanism information should also be helpful to answer diagnostic queries. Observing the presence of a cause's intermediate mechanism variable in addition to the observation of the presence of its effect should make it even more likely that a cause candidate is present. No studies have so far been conducted that test this. Furthermore, it would be interesting to run experiments that directly compare judgments of the diagnostic probability and the probability of singular causation. As we have mentioned, the diagnostic probability of the presence of a candidate cause given the presence of its effect and the probability that a candidate cause actually is the singular cause of the observed effect are not the same (see also Meder et al., 2014). For example, the probability that a person with lung cancer (e) is a smoker (c), $P(c|e)$, is not the same as the probability that it actually was the smoking that caused this person's lung cancer, $P(c \rightarrow e|c, e)$. In some smokers with lung cancer, the singular cause of their disease might be the exposure to asbestos, for example. To give an example using our test cases from Experiment 1, consider test case 22 from segment 4. This is a test case in which neither the target cause nor its mechanism variables

[6] The only possibility of how $P(c \rightarrow e|c, e)$ could be 1.0 in this case is to assign a value of 0 to our model's α parameter. $\alpha = 0$ models a situation in which we rule out that the target cause is causally preempted by an alternative cause. If alternative causes are unobserved, preemption cannot be ruled out, however, and α must be >0.

are observed, but it is observed that the competing castle's telegraph tower forwarded a message to the palace. The probability that the target castle is the singular cause of the alarm in the palace is very low for this test case. By contrast, given the parameterization of the causal structure that we used (see Fig. 2), the diagnostic probability that the target castle sent a message is much higher (see our supplementary file at https://osf.io/3bwyv/) because the lower boundary for the diagnostic probability is the target cause's base rate (see also Meder et al., 2014). As another example, imagine a genetic mutation that is present in almost all individuals of a population but that only rarely causes a specific symptom. The same symptom is often caused by some alternative cause, though. In this population, the probability that someone with the symptom also has the genetic mutation is at least as high as the mutation's base rate. The probability that the mutation actually caused the person's symptom is lower, however, because the mutation has a low causal strength and because there exists an alternative cause that often produces the same symptom. In general, a suitable test condition for future studies asking subjects either diagnostic probability or singular causation queries is one in which a candidate cause has a very high base rate but very low causal strength, and in which alternative causes of the effect are frequent and strong. In such a context, the diagnostic probability of the cause's presence should be high, but the probability that it actually caused the effect in a given case should be low.

8.2. Conclusion

The present research showed both formally and empirically that information about causal mechanisms is an important source of information when it comes to the assessment of singular causation. We have shown that a number of factors informed by mechanism information, including the structure and strength of causal relations, affect singular causation judgments. Future studies need to further investigate boundary conditions of people's competency to assess singular causation relations.

Declaration of Competing Interest

None.

Acknowledgments

We would like to thank Sarah Placì and Ralf Mayrhofer for helpful comments.

This research was supported by a research grant of the Deutsche Forschungsgemeinschaft (WA 621/24-1).

References

Ahn, W., & Bailenson, J. (1996). Causal attribution as a search for underlying mechanisms: An explanation of the conjunction fallacy and the discounting principle. *Cognitive Psychology, 31*, 82–123.

Ahn, W., Kalish, C. W., Medin, D. L., & Gelman, S. A. (1995). The role of covariation versus mechanism information in causal attribution. *Cognition, 54*, 299–352.

Bird, A. (2005). Abductive knowledge and holmesian inference. In T. S. Gendler, & J. Hawthorne (Eds.), *Oxford studies in epistemology* (pp. 1–31). Oxford: Oxford University Press.

Bird, A. (2007). Inference to the only explanation. *Philosophy and Phenomenological Research, 74*, 424–432.

Bird, A. (2010). Eliminative abduction: Examples from medicine. *Studies in History and Philosophy of Science Part A, 41*, 345–352.

Buchanan, D. W., & Sobel, D. M. (2011). Mechanism-based causal reasoning in young children. *Child Development, 82*, 2053–2066.

Buehner, M. J. (2017). Sapce, time, and causality. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 549–564). New York: Oxford University Press.

Cartwright, N. (1989). *Nature's capacities and their measurement*. Oxford: Clarendon Press.

Cartwright, N. (2015). Single case causes: What is evidence and why. In J. Reiss (Ed.), *Philosophy of science in practice*. Dordrecht: Springer.

Cartwright, N. (2017). *How to learn about causes in the single case*. CHESS Working Paper no 2017-2004. https://www.dur.ac.uk/resources/chess/CHESSK4UWP_2017_04_Cartwright.pdf.

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

Cheng, P. W., & Buehner, M. J. (2012). Causal learning. In K. J. Holyoak (Ed.), *The Oxford handbook of thinking and reasoning* (pp. 210–233). New York: Oxford University Press.

Cheng, P. W., & Lu, H. (2017). Causal invariance as an essential constraint for creating a causal representation of the world: Generalizing the invariance of causal power. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 65–84). New York: Oxford University Press.

Cheng, P. W., & Novick, L. R. (2005). Constraints and nonconstraints in causal learning: Reply to White (2005) and to Luhmann and Ahn (2005). *Psychological Review, 112*, 694–706.

Cimpian, A., & Erickson, L. C. (2012). The effect of generic statements on children's causal attributions: Questions of mechanism. *Developmental Psychology, 48*, 159–170.

Danks, D. (2005). The supposed competition between theories of human causal inference. *Philosophical Psychology, 18*, 259–272.

Danks, D. (2017). Singular causation. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 201–215). New York: Oxford University Press.

Fernbach, P. M., Darlow, A., & Sloman, S. A. (2011). Asymmetries in predictive and diagnostic reasoning. *Journal of Experimental Psychology: General, 140*(2), 168–185.

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

Glennan, S. (2017). *The new mechanical philosophy*. Oxford, UK: Oxford University Press.

Glymour, C. (2001). *The mind's arrows: Bayes nets and graphical causal models in psychology*. Cambridge, MA: MIT press.

Glymour, C. (2003). Learning, prediction and causal bayes nets. *Trends in Cognitive Sciences, 7*, 43–48.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., & Danks, D. (2004). A theory of causal learning in children: Causal maps and bayes nets. *Psychological Review, 111*, 3–32.

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

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

Halpern, J. Y., & Pearl, J. (2005). Causes and explanations: A structural-model approach. Part I: Causes. *The British Journal for the Philosophy of Science, 56*, 843–887.

Hart, H. L. A., & Honoré, T. (1985). *Causation in the law*. Oxford, UK: Oxford University Press.

Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences, 8*(6), 280–285.

Hitchcock, C. (2007). Prevention, preemption, and the principle of sufficient reason. *The Philosophical Review, 116*, 495–532.

Hitchcock, C. (2009). Causal modelling. In H. Beebee, C. Hitchcock, & P. Menzies (Eds.), *The oxford handbook of causation* (pp. 299–314). New York: Oxford University Press.

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

Johnson, S. G., & Ahn, W. (2017). Causal mechanisms. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 127–146). New York: Oxford University Press.

Johnson, S. G., & Ahn, W.-k. (2015). Causal networks or causal islands?. the representation of mechanisms and the transitivity of causal judgment. *Cognitive Science, 39*, 1468–1503.

Johnson, S. G., & Keil, F. C. (2018). Statistical and mechanistic information in evaluating causal claims. In T. T. Rogers, M. Rau, X. Zhu, & C. W. Kalish (Eds.), *Proceedings of the 40th annual conference of the cognitive science society* (pp. 618–623). Austin, TX: Cognitive Science Society.

Lagnado, D. A., & Gerstenberg, T. (2017). Causation in legal and moral reasoning. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 565–601). New York: Oxford University Press.

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

Lagnado, D. A., & Speekenbrink, M. (2010 01). The influence of delays in real-time causal learning. *The Open Psychology Journal, 3*, 184–195.

Liljeholm, M., & Cheng, P. W. (2007). When is a cause the "same"?. coherent generalization across contexts. *Psychological Science, 18*, 1014–1021.

Lipton, P. (2004). *Inference to the best explanation*. London: Routledge.

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

Lombrozo, T., & Vasilyeva, N. (2017). Causal explanation. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 415–432). New York: Oxford University Press.

Lu, H., Yuille, A. L., Liljeholm, M., Cheng, P. W., & Holyoak, K. J. (2008). Bayesian generic priors for causal learning. *Psychological Review, 115*, 955–982.

Meder, B., Hagmayer, Y., & Waldmann, M. R. (2009). The role of learning data in causal reasoning about observations and interventions. *Memory & Cognition, 37*(3), 249–264.

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

Meder, B., & Mayrhofer, R. (2017b). Singular causation. In M. R. Waldmann (Ed.), *The Oxford handbook of causal reasoning* (pp. 433–458). New York: Oxford University Press.

Meder, B., Mayrhofer, R., & Waldmann, M. R. (2014). Structure induction in diagnostic causal reasoning. *Psychological Review, 121*, 277–301.

Novick, L. R., & Cheng, P. W. (2004). Assessing interactive causal influence. *Psychological Review, 111*, 455–485.

Park, J., & Sloman, S. A. (2013). Mechanistic beliefs determine adherence to the markov property in causal reasoning. *Cognitive Psychology, 67*, 186–216.

Paul, L. A., & Hall, E. J. (2013). *Causation: A user's guide*. New York: Oxford University Press.

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

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

Rehder, B., & Waldmann, M. R. (2017). Failures of explaining away and screening off in described versus experienced causal learning scenarios. *Memory & Cognition, 45*, 245–260.

Rottman, B. M., & Hastie, R. (2014). Reasoning about causal relationships: Inferences on causal networks. *Psychological bulletin, 140*, 109–139.

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

Russo, F., & Williamson, J. (2011). Generic versus single-case causality: The case of autopsy. *European Journal for Philosophy of Science, 1*, 47–69.

Sloman, S. (2005). *Causal models: How people think about the world and its alternatives*. New York: Oxford University Press.

Stephan, S., Mayrhofer, R., & Waldmann, M. R. (2018). Assessing singular causation: The role of causal latencies. In T. Rogers, M. Rau, X. Zhu, & C. Kalish (Eds.), *Proceedings of the 40th annual conference of the cognitive science society* (pp. 1080–1085). Austin, TX: Cognitive Science Society.

Stephan, S., Mayrhofer, R., & Waldmann, M. R. (2020). Time and singular causation - A computational model. *Cognitive Science, 44*, e12871.

Stephan, S., Tentori, K., Pighin, S., & Waldmann, M. R. (2021). Interpolating causal mechanisms: The paradox of knowing more. *Journal of Experimental Psychology: General*.

Stephan, S., & Waldmann, M. R. (2018). Preemption in singular causation judgments: A computational model. *Topics in Cognitive Science, 10*, 242–257.

Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*, 53.

Waldmann, M. R. (2001). Predictive versus diagnostic causal learning: Evidence from an overshadowing paradigm. *Psychonomic Bulletin & Review, 8*, 600–608.

Waldmann, M. R. (Ed.). (2017). *The Oxford handbook of causal reasoning*. New York: Oxford University Press.

Waldmann, M. R., Cheng, P. W., Hagmayer, Y., & Blaisdell, A. P. (2008a). Causal learning in rats and humans: a minimal rational model. In N. Chater, & M. Oaksford (Eds.), *Prospects for bayesian cognitive science* (pp. 453–484). Oxford, UK: Oxford University Press.

Waldmann, M. R., Cheng, P. W., Hagmayer, Y., & Blaisdell, A. P. (2008b). Causal learning in rats and humans: A minimal rational model. In N. Chater, & M. Oaksford (Eds.), *The probabilistic mind. prospects for bayesian cognitive science* (pp. 453–484). Oxford, UK: Oxford University Press.

Waldmann, M. R., & Hagmayer, Y. (2001). Estimating causal strength: The role of structural knowledge and processing effort. *Cognition, 82*, 27–58.

White, P. A. (1989). A theory of causal processing. *British Journal of Psychology, 80*(4), 431–454.

Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General, 136*, 82–111.

Woodward, J. (2011). Mechanisms revisited. *Synthese, 183*(3), 409–427.