COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 39 (2015) 65–95

ISSN: 0364-0213 print / 1551-6709 online
DOI: 10.1111/cogs.12132

Agents and Causes: Dispositional Intuitions As a Guide to Causal Structure

Ralf Mayrhofer, Michael R. Waldmann

Department of Psychology, University of Göttingen

Received 10 September 2012; received in revised form 23 September 2013; accepted 27 September 2013

Abstract

Currently, two frameworks of causal reasoning compete: Whereas dependency theories focus on dependencies between causes and effects, dispositional theories model causation as an interaction between agents and patients endowed with intrinsic dispositions. One important finding providing a bridge between these two frameworks is that failures of causes to generate their effects tend to be differentially attributed to agents and patients regardless of their location on either the cause or the effect side. To model different types of error attribution, we augmented a causal Bayes net model with separate error sources for causes and effects. In several experiments, we tested this new model using the size of Markov violations as the empirical indicator of differential assumptions about the sources of error. As predicted by the model, the size of Markov violations was influenced by the location of the agents and was moderated by the causal structure and the type of causal variables.

Keywords: Causal reasoning; Causal dispositions; Agency; Markov condition; Causal Bayes nets

1. Introduction

Research on causal reasoning has often been organized around dichotomies. An important recent debate has centered on the differences between views focusing on the covariation between causes and effects and views focusing on mechanisms (see Waldmann & Hagmayer, 2013; for an overview). Although there have been various attempts to reconcile the covariation and mechanism views under an integrated theory, for example, causal Bayes nets (Gopnik et al., 2004; Sloman, 2005; Waldmann, 1996), there are versions of

Correspondence should be sent to Ralf Mayrhofer, Department of Psychology, University of Göttingen, Gosslerstr. 14, 37073 Göttingen, Germany. E-mail: rmayrho@uni-goettingen.de

Portions of this research were presented in the meetings of the Psychonomic Society in 2007, the Annual Meeting of the Society of Mathematical Psychology in 2008, and the Annual Meeting of the Cognitive Science Society in 2010 (Mayrhofer, Hagmayer, & Waldmann, 2010).

mechanism theories that have resisted an easy integration under covariation theories. In particular, adherents of dispositional theories of causation (e.g., force dynamics) have argued that these theories contradict covariation-based theories (e.g., Mumford & Anjum, 2011; Pinker, 2007; White, 2006, 2009; Wolff, 2007). In the present article, we will focus on causal Bayes nets and dispositional theories of causation as two apparently incompatible, independently pursued views of causal representation. We are going to explore whether a more fruitful approach is to view these two frameworks as complementary.

1.1. Two frameworks of causal reasoning

1.1.1. The dependency view: Causes, effects, and their dependencies

The dependency view of causation, which has been stimulated by Hume's (1748/1977) theory, underlies several psychological theories that otherwise greatly differ—including associative theories (see López & Shanks, 2008), covariation theories (e.g., Cheng & Novick, 1992), power PC theory (Cheng, 1997), causal model theories (e.g., Gopnik et al., 2004; Mayrhofer & Rothe, 2012; Rehder, 2003a,b; Sloman, 2005; Waldmann & Holyoak, 1992; Waldmann, Holyoak, & Fratianne, 1995), or Bayesian inference theories (Griffiths & Tenenbaum, 2005, 2009; Lu, Yuille, Liljeholm, Cheng, & Holyoak, 2008 Meder, Mayrhofer, & Waldmann, in press; for overviews see Holyoak & Cheng, 2011; Waldmann & Hagmayer, 2013). According to dependency theories, a factor C is a cause of its effect E if E depends upon C. Dependency has been formalized differently in the various philosophical and psychological approaches. Probabilistic theories focus on statistical dependency: Causes raise or lower the probability of their effects (Cheng, 1997; Salmon, 1984). Interventionist theories emphasize the role of interventions that bring about effects (Gopnik et al., 2004; Woodward, 2003). Finally, counterfactual theories describe an event C as a cause of E when it holds that if C had not occurred, E would not have occurred (Lewis, 1973). All these approaches share the view that causes are *difference makers* and that causal relations are fundamentally about dependency.

Currently, the most popular versions of dependency theories are causal model or causal Bayes net theories, which are not restricted to single cause–effect relations but can model complex networks of interconnected causal events (Gopnik et al., 2004; Pearl, 2000; Spirtes, Glymour, & Scheines, 1993; Spohn, 2012; Waldmann & Martignon, 1998). The basic entities of causal Bayes nets are nodes that code variables—such as the presence or absence of events, properties, or facts—and arrows that code causal dependencies between the directly linked variables (for an example, see Fig. 1).

Although the notion of dependency is at the core of all dependency theories, different theories assign different roles to observed dependencies. Whereas older probabilistic and associative theories claim that statistical dependency is all there is to causality, more recent accounts try to bridge the gap between statistical dependency and mechanism knowledge. According to these approaches, the observed statistical dependencies (i.e., covariations) are indicators of hidden causal powers and mechanisms (e.g., Cheng, 1997). Pearl (2000) aptly called causal arrows *mechanism placeholders*.

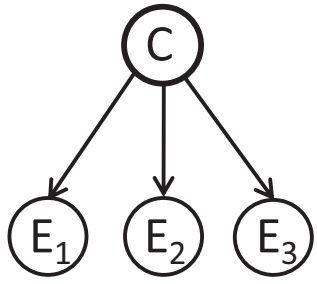

Fig. 1. An example of a simple common-cause structure with a cause variable C and three effect variables, E_1, E_2, and E_3.

Various psychological experiments have lent support to the view that everyday causal reasoning can be fruitfully modeled as reasoning with causal Bayes nets (see Rottman & Hastie, 2014; Waldmann & Hagmayer, 2013; for overviews). However, despite many successful applications of dependency theories, in particular causal Bayes nets, some recent results have cast doubts on the plausibility of some of the key assumptions underlying these models. A number of researchers have started to study the plausibility of the *Markov condition*, which is a central assumption underlying Bayes nets. According to the Markov condition, each variable conditioned upon its direct causes is independent of all other variables except its direct and indirect effects. In philosophy, there has been a debate about the validity of the Markov condition. Cartwright (2007) argued that our causal knowledge is rife with Markov violations and, therefore, concluded that Bayes nets do not adequately express causal knowledge. Hausman and Woodward (1999, 2004), however, claimed that Cartwright's examples of Markov violations can be handled by adding hidden variables to the Bayes net representations. Thus, although a causal model containing only observed variables might exhibit Markov violations, the full model enriched with hidden variables honors the Markov condition and explains the observed Markov violation. The computational model we will introduce below will demonstrate this property.

A number of psychological studies have also shown that people tend to violate the Markov condition. For example, in some experiments of Rehder and Burnett (2005), subjects were presented with common-cause models (see Fig. 1) and were asked to rate the conditional probability of one of the effects given the state of its cause C. The crucial manipulation was whether other collateral effects of C were also present or absent. According to the Markov condition, participants' ratings should be invariant across these conditions. Contrary to this prediction, the ratings were clearly sensitive to the states of the other effects of C. The more collateral effects were present, the higher the ratings of the conditional probability of the target effect given the presence of C. Thus, the results indicated that subjects did not represent the instructed common-cause model in the intended way (as displayed in Fig. 1).

Similar results were obtained by Walsh and Sloman (2007): In one of their experiments, for example, they presented a cover story in which C represents the playing of

loud music in an apartment building, E_1 the effect that the left neighbors complain, and E_2 the effect that the right neighbors complain. As in Rehder and Burnett's (2005) studies, the target question was about the likelihood of, for example, E_2 given that C is present. Without any further explanations, there was a tendency to violate the Markov condition: The likelihood of E_2 given C was rated higher when E_1 was present than when it was absent. Walsh and Sloman explained this finding by assuming that subjects tried to generate explanations of the unexpected pattern with a present cause and one missing effect. They argued that there was a tendency to explain such a pattern with a common disabling condition that affects both effects. For example, subjects might have explained the fact that the left neighbors did not complain by assuming that the tenant who was playing the music had invited all neighbors to a party. This common disabling condition would explain the Markov violation. Walsh and Sloman tested this assumption by running conditions in which they suggested disablers that were restricted to only one effect (e.g., the left neighbors did not complain because they were not staying in their apartment building at this time). In this case, it was observed that the likelihood ratings regarding the right neighbors were less affected by the status of the left neighbors. Thus, the previously observed Markov violation was diminished.

Although it has been demonstrated that Markov violations of this kind are the rule rather than the exception in causal inference tasks, none of the psychological researchers has questioned the general usefulness of Bayes nets as a tool for modeling causal cognition (see Rottman & Hastie, 2014). Therefore, in line with the proposal of Hausman and Woodward (1999, 2004), the researchers augmented the causal Bayes net representation. However, the empirical results also make clear that the usual assumption that people directly adopt the instructed causal model needs to be revised. The results suggest that people bring world knowledge to the task that suggests the addition of hidden variables, which would explain the apparent Markov violation.

It is fairly plausible to assume that with real-world scenarios experimental subjects do not always stick to the instructions provided in the cover stories but tend to augment the instructed model with additional hidden variables coding their background knowledge about mechanisms. In some cases, the instructed cover stories draw from elaborate prior knowledge suggesting fairly specific additional variables (e.g., Walsh & Sloman, 2007; see also Park & Sloman, 2013). There are other cases, however, in which it is less clear where additional background assumptions come from. For example, Rehder and Burnett (2005) found similar patterns of Markov violations as in the real-world cases with tasks in which only abstract causal variables were used (labeled A, B, C, and D). In fact, it has proven nearly impossible to get inferences in the Rehder/Burnett paradigm that do not exhibit Markov violations.

These findings create a puzzle. Why should an abstract cover story that simply labels the variables with letters and does not mention additional hidden variables generate similar Markov violations as real-world scenarios in which possible common disablers are explicitly mentioned in the instruction or can easily be recruited from background knowledge? Would it not be plausible, at least in the case of abstract cover stories, that people accept, for example, the common-cause model that has been presented to them? We will return to this puzzle below.

1.1.2. The dispositional view: Agents, patients, and force dynamics

A completely different view answers the question why an observed lawfulness holds by focusing on the participants involved in a causal interaction, for example, Ball A and Ball B in Michotte's (1963) task, or aspirin and a person with a headache in a medical scenario. *Dispositional* theories of causation would say, for example, that the ingestion of Aspirin relieves headaches because aspirin has an intrinsic property, a disposition (or capacity, potentiality, power) to relieve headaches in suitable organisms, that interacts with the disposition of human bodies to be influenced by aspirin. According to this view, dependency relations are secondary; they arise as a product of the interplay of causal participants that are endowed with specific dispositions.

There is a huge philosophical and linguistic literature devoted to the question of how dispositions should be semantically analyzed and how dispositional ascriptions relate to categorical properties of objects or substances, which we cannot discuss here (see, for example, Kistler & Gnassounou, 2007; Mumford, 2003; Mumford & Anjum, 2011; Spohn, 2012; Talmy, 1988). In psychology, which is our focus, one variant of dispositional theories, *force dynamics*, has become increasingly popular in recent years. Pinker (2007) argued that force dynamics is a major competitor of Bayes net theories because it allows us to model intuitions about the generational processes underlying observed covariations. Force dynamics has been initially developed and empirically tested in linguistics in the context of verb semantics (see Levin & Rappaport Hovav, 2005; Riemer, 2010; Talmy, 1988). Its concepts can be traced back to Aristotle, who explained efficient causation as a consequence of the interaction of two entities, an agent and a patient. An agent is, according to Aristotle, a substance operating on another substance, the patient, which is passive with respect to the process of operation and sometimes resists the agent's influence. The agent, which acts upon the patient, therefore, has the *disposition*, *capacity*, or *power* to act; and the patient has the disposition to be acted upon.

Within psychology, White (2006, 2009), who focuses on perceptual causation, and Wolff (2007; see also Wolff, Barbey, & Hausknecht, 2010) have adopted and developed variants of force dynamics. Because we do not investigate perceptual causation in this article, we concentrate here on Wolff's theory. The primary aim of Wolff's force theory is to elucidate our understanding of abstract causal concepts, such as *cause*, *prevent*, *enable*, and *despite* (see also Wolff, 2012; Wolff & Song, 2003; Wolff et al., 2010). Force theory states that people evaluate configurations of forces attached to agents (affectors in Wolff's terminology) and patients with respect to an endstate. Forces can be physical, psychological (e.g., intentions), or social (e.g., peer pressure). Agent–patient interactions, then, are analyzed in terms of three components: (a) the prior tendency of a patient toward the endstate, (b) the concordance between agent and patient, and (c) whether the endstate is reached or not. The states of these three components determine people's conceptualization of causal scenarios. For instance, the sentence "The wind moved the boat to the harbor" would be construed as an example of *cause* because the agent (wind) exerts a force on the patient (boat) that overcomes the patient's prior tendency (standing still) and because the patient reaches the endstate (harbor). Similar analyses can be offered for other abstract causal concepts, such as *prevent*, *enable*, or *despite*.

One limitation of the current implementation of force theory is that it is restricted to modeling semantic intuitions about interactions between one agent and one patient (but see Wolff et al., 2010, for an extension to chains of interactions). Obviously, people can also reason about more complex causal models, such as common-cause or common-effect models. Moreover, causal knowledge allows us to make various specific inferences that are presently not captured by force dynamics: We form expectations about future events and explanations of present events (i.e., predictive and diagnostic inferences), or think about what might have happened (i.e., counterfactual inferences) and about people's accountability (i.e., blame/responsibility attributions).

1.2. The interaction of dispositional intuitions and dependency models

Although dependency models and dispositional theories have been claimed to be competitors, our brief review has demonstrated that both accounts may have weaknesses and strengths that seem complementary. Both frameworks address different aspects of causal relations and different *causal relata*. Dependency theories connect variables that refer to the presence and absence of events or facts. By contrast, dispositional theories start with an analysis of the objects that are potentially involved in causal interactions. The objects may have dispositional properties that, when placed in the right context, can give rise to the causal dependencies between events modeled by dependency theories. For example, an aspirin pill in a bottle is not a cause of anything—but due to its chemical properties, it has the potential of curing headache when swallowed by a patient having the right physiological properties guaranteeing the drug's success. Whereas dependency theories might model this situation as the presence of the event "swallowing aspirin" (i.e., cause) leading to a "relieve of headache" (i.e., effect), dispositional theories start with analyzing the characteristics of the causal participants (e.g., aspirin pills, the human body), which have the potential to generate a specific observable causal relation between events.

The strict separation between dispositional theories and dependency theories in the literature has prevented researchers from thinking about possible interactions between these two frameworks. For example, research within the dependency framework has largely neglected the powerful role of linguistic instructions conveyed in the cover stories. Typically, researchers fiddled with their cover story until it served the intended goal while they forgot that there is actually an extensive literature about the semantics of causal verbs and other linguistic expressions used to convey causal intuitions.

Our main claim in the present article is that abstract dispositional intuitions (e.g., about agency), in part conveyed by linguistic expressions, may guide the formation of causal models when specific world knowledge about mechanisms is not available or vague. The link between the formation of causal models and dispositional intuitions may also explain why Rehder and Burnett (2005) found Markov violations even with abstract scenarios. Our semantic intuitions about causal expressions may bias us toward augmenting the instructed causal model even when only abstract descriptions of causal relations are provided.

One attractive feature of dispositional theories and force dynamics in particular is that these theories are capable of expressing abstract intuitions about mechanisms without necessarily requiring specific mechanism knowledge. World knowledge will, of course, be used if it is available (as, for instance, in Walsh & Sloman, 2007). But sometimes, we may only understand that one entity is an agent and the other a patient without having further knowledge about the mechanisms connecting these entities. Recent studies about mechanism knowledge showed that subjects, unless they are domain experts, are typically quite naïve about mechanisms, even for objects they encounter every day (e.g., cars, TV; see Keil, 2012; Rozenblit & Keil, 2002). Consequently, force dynamics and related theories do generally not model scientific knowledge but rather vague intuitions about abstract properties of causal objects. These intuitions may even contradict physical theories, such as Newtonian mechanics (White, 2009).

1.2.1. Agents and causes: Two distinguishable causal concepts

Within dependency theories, a *cause* is an event (or fact) whose presence makes a difference with respect to the presence of another event (or fact), the *effect*. Thus, the effect variable is dependent on the cause variable. By contrast, dispositional theories distinguish between *agents* and *patients*, which refer to the causal participants (i.e., objects or entities) in a causal interaction. In the majority of cases, causes describe events in which agents are involved, and effects refer to events in which patients take part (e.g., White, 2006). Not surprisingly, cover stories used in the literature often describe scenarios in which causes and agents are confounded, for example: medicine causing headache (e.g., Buehner, Cheng, & Clifford, 2003), food causing allergies (e.g., Shanks & Darby, 1998), chemicals or radiation causing the expression of genes or diseases (e.g., Griffiths & Tenenbaum, 2005; Perales, Catena, & Maldonado, 2004), or fertilizers causing plants to bloom (e.g., Lien & Cheng, 2000). Although all these cover stories were used to test dependency models, the putative cause (e.g., medicine, food, chemicals, fertilizer) is actually an object (or a substance)—the agent—serving as a semantic shortcut to the relevant cause event: the agent's interaction with an (often implicit) patient.

Although cause events typically involve agents and effect events patients, there are situations in which this mapping is reversed (see Fig. 2). For example, consider a situation in which a pedestrian stops when facing a red traffic light. In this case, many people will view the pedestrian as the agent: She perceives the red light and then decides to stop. However, an outside observer focusing on dependency relations sees a situation in which the light controls the behavior of the pedestrian: Intervening in the traffic light changes the behavior of the pedestrian, but intervening in the pedestrian does not influence the traffic light. Clearly, the dependency relation goes from traffic light (cause) to the stopping of the pedestrian (effect), although the pedestrian is conceived as the agent in this situation.

In (psycho) linguistics, these kinds of reversals have been studied in the context of the semantic analysis of *psychological verbs* (psych verbs; see Brown & Fish, 1983; Landau, 2010; Pinker, 1991; Rudolph & Försterling, 1997; Semin & Fiedler, 1988, 1991). Psych verbs present a puzzle for linguistics because they often reverse the usual mappings between causal roles and grammatical categories (see Landau, 2010). In "Peter frightens

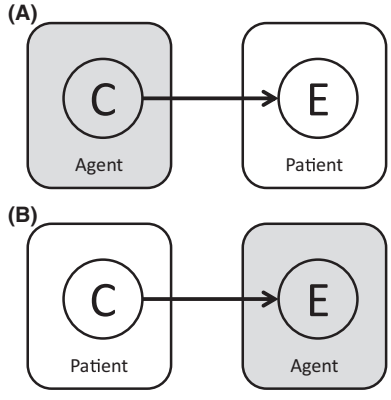

Fig. 2. A single cause–effect relation with (A) the agent role attached to the cause event (C) and (B) the agent role attached to the effect event (E).

Mary," Peter, the subject of the sentence, plays the role of the agent and Mary the role of the patient. But in the context of psych verbs, the agent can be located in the grammatical accusative object role (e.g., "Mary fears Peter"). For the present purposes, the crucial feature of psych verbs is that they often describe scenarios in which the agent is involved in an effect-related event. In, for example, "Bill listens to Joe," Bill, the experiencer, plays the active role (i.e., agent). However, within a causal dependency model, Bill's acquired knowledge would be represented as an effect of what Joe had said. Joe's speech is part of the cause event because it is happening temporally prior to Bill's acquisition of knowledge and is its necessary precondition. Other examples of such verbs, which do not necessarily have to refer to an animate agent, are *perceive*, *read*, *receive*, and *detect*.

So far, we have described the agent and patient roles as belonging either to the cause or to the effect side. However, the psycholinguistic literature has shown that the assignment is not always clear-cut. For example, in a situation in which a teacher tries to convey knowledge to her pupils, both sides may be viewed as partially active. Both the teacher and the student need to actively intend the successful outcome of the transaction. Dowty (1991) proposed that the agent and patient classification is not based on definitional criteria; agency rather is a family resemblance concept (see also Mayrhofer & Waldmann, 2013, for examples involving perceptual causality). According to Dowty's theory, both the agent and the patient roles are assigned based on a number of criteria. In Dowty's theory, the *agent* features include (among others) (a) volitional involvement in the event or state, (b) sentience (and/or perception), and (c) causing an event or change of state in another participant. The *patient* features include (among others) (a) undergoes change of state, and (b) causally affected by another participant. This theory also allows to model ambiguous cases in which agency is distributed across different participants.

1.2.2. The role of responsibility and error attribution in causal inferences

So far, we have argued that agents and causes are distinguishable concepts. In the following section, we will show how agency intuitions may interact with dependency

knowledge and how this interaction can help us explain Markov violations that pose a problem for causal Bayes net theories.

In causal relations, we typically expect that the effect will happen when the cause is present (see Goldvarg & Johnson-Laird, 2001; Mayrhofer & Waldmann, 2011). This is a case of a successful instantiation of a causal relation in which the agent succeeds in producing the effect when interacting with the patient. However, there are also unsuccessful instantiations in which the cause is present but the effect is absent. In this case, the agent's interaction with the patient does not generate the to-be-expected result. Settings in which violations of the Markov condition can be observed generally involve such unsuccessful instantiations of causal relations: As discussed above, expectations about a target effect's presence (given the cause is present) usually are lowered, the more other effects of the same cause are observed to be absent (Rehder & Burnett, 2005; Walsh & Sloman, 2007; see also Mayrhofer, Goodman, Waldmann, & Tenenbaum, 2008). Interestingly, in all the cover stories used to test whether subjects honor the Markov condition, the agents were part of the cause event.

Our key empirical hypothesis is that the size of Markov violations is moderated by the location of the agents and patients within the causal dependency relations. We expect larger Markov violations in common-cause models when an agent is involved in the cause event (as in previous research) than when it is involved in the effect events. The main reason for this prediction is that in cases in which causes fail to generate their effects, responsibility for this failure is differentially attributed to the agent and the patients involved in the causal relations. How, then, is responsibility assigned in successful and unsuccessful causal transmissions?

For the successful case, previous research indicates that agents are typically held more responsible for the success than patients. For example, White (2006) looked at cases in perceptual scenarios in which agents were involved in the cause event and patients in the effect event. He found what he called "causal asymmetry": The agent's contribution to the success (i.e., effect is present) is typically overestimated relative to the contribution of the patient. For instance, in Michotte-like physical interactions between two balls, more force is attributed to the launching than to the launched ball (White, 2007). Additional evidence comes from a number of studies on moral reasoning: For example, Cushman (2008) showed that human agents tend to be blamed for negative outcomes when there is a strong causal relation between act and outcome. However, introducing several agents causing the outcome weakens responsibility attributions to the agent (Gerstenberg & Lagnado, 2010; Lagnado & Channon, 2008; Woolfolk, Doris, & Darley, 2006).

How about the unsuccessful cases? Responsibility attributions in cases of causal failure are more ambiguous: In a situation in which an agent interacts with a patient, either the agent force turns out to be too weak to affect the patient, or the resistance of the patient is too strong for the agent. As in the case of a successful causal interaction, responsibility attribution is here a consequence of a relational assessment. It depends on assumptions about the variability of the agent relative to the patient force in a set of causal events. In the absence of prior knowledge about the causal scenario,

we expect that agents, which tend to be held accountable for successful causal interactions, are also primarily held accountable for the occasional failure. But obviously, these intuitions might also be influenced by real-world knowledge. Moreover, the possibility of distributing agency across different causal participants, which we mentioned in the context of Dowty's (1991) theory, also may influence responsibility assessments. In this case, both interacting participants might be held responsible for the failed causal transmission.

How can responsibility attributions explain Markov violations in human causal inferences? As described above, subjects in experiments studying Markov violations are typically requested to judge the likelihood of a target effect's presence when its cause is present while the presence or absence of the other collateral effects is systematically manipulated (i.e., common-cause model; see Fig. 1). According to the Markov condition, the rating should only depend upon the state of the cause but should be independent of the states of the collateral effects. However, the ratings typically decrease the more collateral effects are absent. If in such a common-cause model the cause event is associated with the agent (as in previous research), the agent will mainly be held responsible for these unsuccessful outcomes. This should lead to the expectation that all effect events dependent on the same agent should also be affected by its current deficit. Under the assumption that the patients are of the same type, we expect that all effect events should be affected by the agent's deficit to an equal extent. Thus, lowered ratings and, therefore, a substantial Markov violation is predicted in this case. However, if the cover story or subjects' prior knowledge suggests different types of patients, a more complex version of the model incorporating category knowledge can be developed (see Mayrhofer et al., 2008, for such a case).

By contrast, if the effect events are instead primarily viewed as involving different agents (interacting with the patient involved in the cause event), a failure within a specific cause–effect relation should less likely be generalized to the other cause–effect relations that involve different, independently operating agents. Unless there is a reason to assume a hidden common cause affecting all operating agents at once, it seems plausible that failures of individual agents are caused by unknown internal mechanisms that do not affect the workings of other agents. This assumption of independence of the agents should lead to a sizable reduction of the Markov violation.

1.3. A Bayes net account of error attribution

One of our main goals is to show how dispositional intuitions about agents and patients can be mapped onto causal model representations that allow reasoners to make inferences within complex networks. We have implemented our hypothesis of differential error attribution to agents versus patients using a causal Bayes net representation (see also Rehder & Burnett, 2005; for a related model). In the case of Markov violations, our key claim is that differential assignments of agents and patients to the cause and effect events lead to different attributions of responsibility for observed failure (i.e., cause is present but effect is absent). Standard Bayes nets are ill equipped to express

these differential responsibility attributions because they typically collapse all the sources of errors for each cause–effect pair into the single corresponding causal strength parameter. The causal strength parameter, w_C, expresses the probability of the cause producing its effect (in the hypothetical absence of alternative causes; see Cheng, 1997; Griffiths & Tenenbaum, 2005) and, therefore, explains why in the nondeterministic case (i.e., $w_C < 1$) occasionally the cause co-occurs with the absence of its effect. Causal strength by itself is not sensitive to the location of the error. In complex causal Bayes nets, each causal relation has its own causal strength parameter, and all links are thought to be independent of each other (due to the Markov condition). Thus, if something goes wrong in one relation, this failure will not affect the other relations involving the same cause event.

To account for differential error attributions, we conceptually augmented the network by separating two sources of failure: one tied to the participant involved in the cause event (F_C) and one tied to the participant involved in the effect event (F_E; see Fig. 3A). We propose, therefore, that each cause C contains an independent hidden *preventive* node F_C that is connected to all of the cause's effects (see Fig. 3C) and that can therefore alter the influence of the cause on its different effects at once. This preventive node is an abstract representation of anything that may go wrong with the causal participant involved in the cause event. When the error node is active, it decreases the power of C to bring about its effects. Importantly, the F_C node is linked to *all* effects generated by the

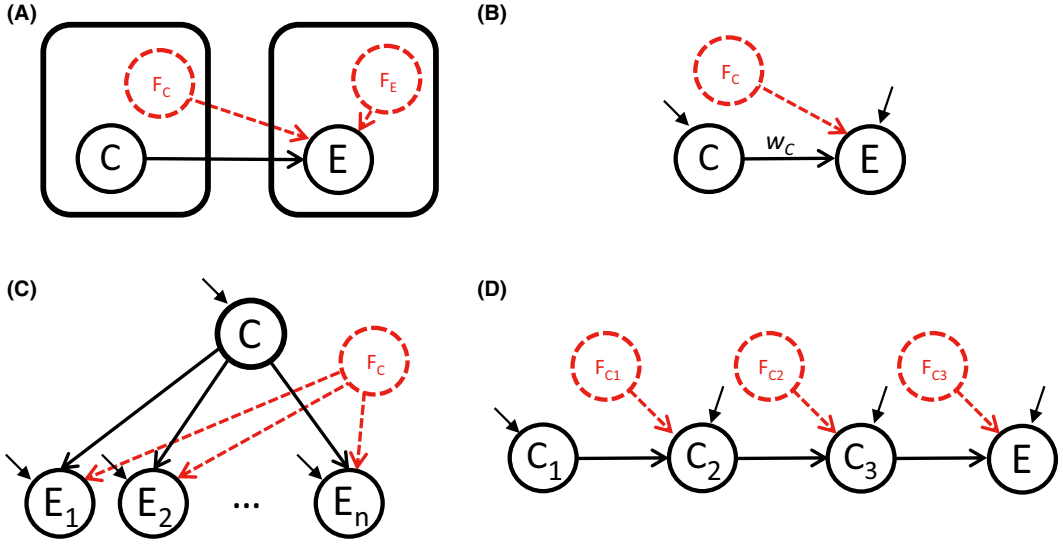

Fig. 3. (A) A causal relation with two sources of failure, a cause-related error node F_C and an effect-related error node F_E; (B) a simplified version with the influences of F_E collapsed into the causal strength parameter w_C; (C) a simple common-cause structure with the common preventive cause-related error node F_C; and (D) an extended causal-chain model with two indirect causes (C_1, C_2), a direct cause (C_3), and a terminal effect (E) with each cause having its own preventive cause-related error node (F_{C1}, F_{C2}, F_{C3}).

cause. Thus, if there is an error on the cause side, all effects of this cause will be affected proportional to the causal strength of the relations that link the F_C node to the different effects. By contrast, the sources of failure on the effect side (F_E) are specific to each effect and, therefore, independent of each other. If anything goes wrong on the effect side, then only the specific effect will be prevented, but not the other effects within the network.

In Fig. 3A, nodes and arrows for both cause-related (F_C) and effect-related (F_E) errors are drawn. However, given that for our simulations we used standard causal models with probabilistic causal strength parameters, we simplified the representation by reserving the probabilistic strength parameters w_C linking causes and effects to express both causal strength and effect errors (F_E; see Fig. 3B). The causal strength parameters (or more specifically, their deviation from being 1), then, summarize all unobserved additional influences that may be caused by unknown external events, by unknown internal mechanisms inherent in the causal participant involved in the effect event, or by other influences that specifically target the effect event. By contrast, the preventive node F_C, being linked to all effects of a target cause, represents all cause-based sources of failure (see Fig. 3C). An alternative representation, which would make similar predictions for our cases, would be a quasideterministic network in which all causal links are deterministic with the base rates of added deterministic inhibitors explaining the observed probabilistic relations (see Buchanan, Tenenbaum, & Sobel, 2010, for such an approach).

To implement our hypothesis of differential error attribution to agents and patients, we manipulated the strength distributions of the different sources of error: The strength of the hidden preventer F_C is (relatively) higher when the agent role is assigned to the cause event than when it is assigned to the effect event. Our error attribution hypothesis only makes relational predictions so that altering the preventive strength of the F_C node suffices to generate different relational patterns. (Note that the model can easily be parameterized to handle cases in which the agent and patient assignments are more ambiguous or in which additional cues shift assessments of the relative strength of the force and resistance of agents and patients.)

In the following sections, we will present our model more formally and will show how the introduction of the common preventive error node F_C affects predictive judgments in different causal structures that we tested in the experiments presented below. We will start with a demonstration that agency assignment does not matter for probabilistic inferences in single cause–effect relations. Then, we are going to apply the model to common-cause and to causal-chain structures to derive empirically testable predictions.

1.3.1. Single cause–effect relations

Given a single cause–effect relation, predictive or diagnostic inferences are not affected by the introduction of the added preventer F_C. As F_C is unobserved, it will be integrated out. For a simple predictive query, that is, inferring the state of the effect E given the state of its cause C, we therefore get the following equations:

$$\begin{aligned} P(E=1|C) &= \sum_{F_C} P(E=1|C,F_C)P(F_C|C) \\ &= \sum_{F_C} P(E=1|C,F_C)P(F_C) \end{aligned} \tag{1}$$

The second step is possible because, in this special case, the inference about the variable F_C does neither depend on C (because F_C and C are independent of each other) nor on E (because E is not known; it is the target of the predictive inference). Because only the cause C and its effect E are involved in this inference, this model makes the same predictions as a version without F_C. We, therefore, predict that agency intuitions do not influence predictive reasoning about single causal relations (given that everything else is held constant). This prediction was tested in Experiment 1.

1.3.2. Common-cause models

As discussed above, models with and without F_C do, however, make different predictions in common-cause structures with a single cause C and n effects $E_1, \ldots, E_n$ (see Fig. 1 for an example with $n = 3$). According to the Markov condition, a predictive inference from cause C to its nth effect E_n is independent of all other effects $E_1, \ldots, E_{n-1}$:

$$P(E_n = 1|C, E_1, \ldots, E_{n-1}) = P(E_n = 1|C) \tag{2}$$

By contrast, in our postulated representation of common-cause models, there is one single joint F_C node attached to the effects of the common cause (see Fig. 3C). Thus:

$$\begin{aligned} &P(E_n = 1|C, E_1, \ldots, E_{n-1}) \\ &\quad = \sum_{F_C} P(E_n = 1|C, F_C, E_1, \ldots, E_{n-1})P(F_C|C, E_1, \ldots, E_{n-1}) \\ &\quad = \sum_{F_C} P(E_n = 1|C, F_C)P(F_C|C, E_1, \ldots, E_{n-1}) \end{aligned} \tag{3}$$

The second simplifying step in this derivation is possible because in the network with F_C the Markov condition holds again: Given *C and* F_C, the target effect E_n is independent of its collateral effects. Thus, reasoning within this model can be conceived of as a two-step process: First, based upon the states of the observable variables C and $E_1, \ldots, E_{n-1}$, the state of F_C is inferred; this instantiation of the augmented model is then used to infer E_n.

The crucial difference to the single-link version is that the inference about the preventer F_C does depend on the states of all effects of the respective cause. The model predicts that inferences about the presence of an unobserved target effect E_n are influenced by the number of collateral effects that are present or absent (as this is the crucial manipulation in the paradigm of Rehder & Burnett, 2005). Absent collateral effects lower (via F_C) the ratings for the target effect proportional to their number. This influence is higher

when the agent role is assigned to the cause side (relative strength of F_C is high; see above) than when the agent role is assigned to the effect side (relative strength of F_C is low).

In Fig. 4A, the predictions of the model are shown for a set of causal parameters (base rate of effects 0.33 and causal strength 0.75)[1] and an instantiation of a causal Bayes net in which a binary cause C is either present or absent. We assume that the cause only exerts an influence on its effects when present but not when absent (we call this the 0/1 case), as this is the usual implementation of causal dependency in causal reasoning tasks based upon Cheng's (1997) analysis. When the cause is present (the upper set of lines), the influence of collateral effects is highest in the case of a strong F_C (i.e., agent role assignment to the cause side) in contrast to a weak F_C (i.e., agent role assignment to the effects side). When the cause is absent, however, no such pattern is observed because the effects are only influenced by background factors (i.e., the inference is determined by the base rates of the effects). These predictions were tested in Experiment 2.

The model allows us to make further testable predictions for common-cause structures. Usually, the variables representing causes and effects are binary with the two states coding either absent or present events (0/1 case). In some cover stories, however, the variables are coded as representing two alternative values (i.e., A/B case). Thus, both states of the cause are causally active with respect to the different states of the effect variable (as, e.g., in the cover stories used by Rehder & Burnett, 2005). In this case, two different causal relations are expressed for the different states of the cause variable; both, therefore, should be affected by the location of agents and patients (see Fig. 4B). This prediction was tested in Experiment 3.

Our model accounts for Markov violations by adding a hidden preventive node F_C for each cause. There is an alternative modeling strategy that does not require the addition of nodes, however. Given that the instructions about causal strength are typically vague in

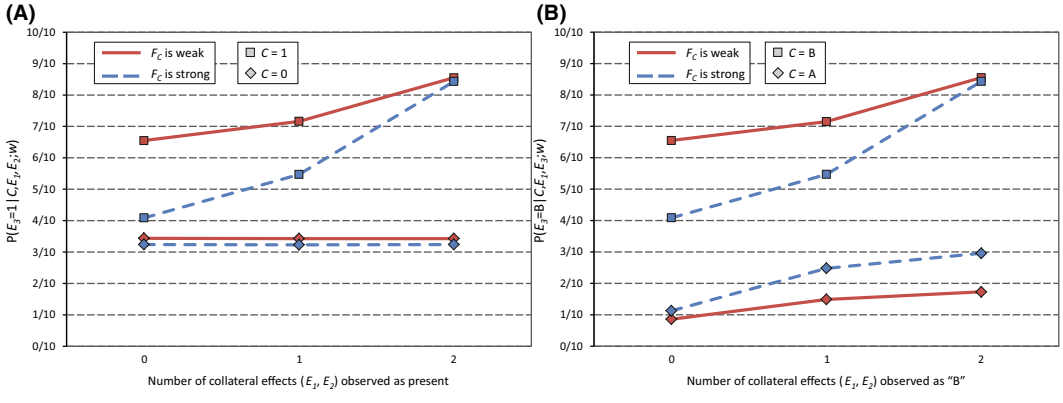

Fig. 4. Model predictions for inferences in a common-cause structure with three effects for (A) a binary cause that is either present or absent (0/1 case) and for (B) a case in which both states of the cause are causally active (A/B case).

the Rehder/Burnett paradigm, it can be argued that the observation of specific causal failures may be viewed as indicative of relatively weak causal strength of a specific cause. This inference might be generalized to the other equivalent effects of the cause in a hierarchical Bayesian model. Thus, such mutual adjustments of causal strength in a common-cause structure may lead to inferences that appear like Markov violations. However, such a model cannot explain the role of the location of agents and patients in predicting the size of Markov violations, the main focus of our research. Therefore, we propose a model without this hierarchical parameter learning mechanism.[2]

1.3.3. Causal-chain models

Our model can be applied to causal chains as well. Causal chains are causal structures that are also very interesting with respect to the Markov condition. In a causal chain, a target effect E is preceded by a chain of causes $C_1, \ldots, C_n$ such that each event only depends on the respective predecessor in the chain. According to the Markov condition, a predictive inference from the cause directly preceding E, C_n, to E does not depend upon the states of the variables in the chain prior to C_n (i.e., the indirect causes of E: $C_1, \ldots, C_{n\text{-}1}$). In our model, preventive nodes are attached to each individual cause (see Fig. 3D). As in the single-link case, the strength of each F_C in the chain does, therefore, not bias people's inferences about the states of the other variables:

$$\begin{aligned} P(E_n = 1|C_1, C_2, \ldots, C_n) &= \sum_{F_{C_1},\ldots,F_{C_n}} P(E_n = 1|C_1, \ldots, C_n, F_{C_1}, \ldots, F_{C_n}) \cdot \\ P(F_{C_1}, \ldots, F_{C_n}|C_1, \ldots, C_n) &= \sum_{F_{C_n}} P(E_n = 1|C_n, F_{C_n}) \cdot P(F_{C_n}) \end{aligned} \tag{4}$$

Thus, when making an inference from a direct cause C_n to its effect E, it should not matter whether the agent role is assigned to C_n or to E. Consequently, our model predicts that in causal-chain structures indirect causes do not affect inferences about the target relation at the end of the chain regardless of the location of the agent or patient within this causal relation—a prediction that was tested in Experiment 4. Of course, this analysis only applies to chains in which there are only direct causal links between adjacent events. More complex temporally ordered structures with indirect or interactive relations or with additional hidden causal events would require a different model.

2. Experimental paradigm

2.1. General cover story

To test our hypothesis that Markov violations are sensitive to the location of agents and patients because of differential error attributions, we chose a scenario in which

agency can be manipulated independently of the dependency relation. We used cover stories adapted from Steyvers, Tenenbaum, Wagenmakers, and Blum (2003), who investigated people's ability to infer causal structure from covariation information. Steyvers and colleagues presented subjects with fictional scenarios about alien mind readers. For example, participants were told that there are three aliens that might be able to read the thoughts of the other aliens; the task, then, was to identify the aliens capable of mind reading given a set of thought configurations. We selected the cover story about mind-reading aliens to minimize real-world knowledge about mechanisms.

To disentangle the roles of causes and agents (see Fig. 2), we modified this cover story. In general, we kept the roles of cause and effect constant across different conditions. One alien, the cause, was described as having a specific thought that causes thoughts in effect aliens. Given that the thought of the cause alien temporally precedes the thoughts of the effect aliens, the direction of the causal arrow within a dependency model is clear. This intuition can also be substantiated by considering what would happen if an external intervention either changed the thought of the cause or of the effect aliens. Only in the former case, the thought of the complementary alien would also be affected.

To manipulate the dispositional roles of causes and effects, we used different verbs describing the underlying causal mechanism. In one condition, the cause alien was described as being capable of *sending* its thoughts (sender condition). This verb should establish the cause alien as the agent and the effect aliens as patients in a successful causal transmission (see Fig. 2A). In this case, we expected that the agent that attempts to transmit its thought into the effect alien should to a substantial extent be held responsible for both success and failure of its attempts. In the contrasting condition, the effect aliens were described as being capable of *reading* the thoughts of the cause alien (reader condition). In this condition, the effect aliens should be viewed as the agents and the cause alien as the patient. A plausible intuition in this case is that the mind reader picks up the thoughts of the cause alien without this alien even being aware of the existence of a causal transmission (and thus exerting no resistance). Thus, this should be a clear case in which the agent (the mind reader), which is located on the effect side, should be held fully responsible for both success and failure of the causal transmissions.

2.2. Causal structures

According to our model, dispositional intuitions about the location of agents and patients guide the structuring and parameterization of a causal Bayes net augmented by hidden preventive nodes. More specifically, the hypothesis is that differential assignments of the agent and patient roles to causal events should, due to differences in error attributions, lead to predictable differences in causal inferences depending upon the causal structures in question. We, therefore, tested our model in five studies that varied the structural properties of the causal scenario, such as the type of causal variables or how the variables are related to each other.

In the first study, we investigated whether intuitions about agency can be dissociated from the cause–effect dependency relation (Experiment 1a) and whether responsibility is

differentially assigned depending on the locations of agent and patient within an otherwise identical causal dependency relation (Experiment 1b). Then, we tested the predictions of our hybrid model for common-cause scenarios with causes that either vary between present and absent (Experiment 2) or between two causally active states (A/B case; Experiment 3). Finally, we tested the prediction that the location of the agent should not matter for predictive inferences within a causal chain (Experiment 4).

3. Experiments 1a and 1b

The aim of Experiments 1a and 1b was to test whether people differentiate between dispositional roles (i.e., agent, patient) and causal dependency (i.e., cause, effect). In many causal scenarios, cause events involve agents so that it may be the case that people routinely conflate agents and causes. In the two experiments, we presented subjects with a single causal relation in which a specific thought (about food) of an alien A tends to be transmitted to alien B. Thus, except for rare cases in which alien B happens to think of food by itself, the thought of alien B is causally dependent on the thought of alien A. From the perspective of dependency theories, the thought of alien A clearly is the cause and the thought of alien B the effect. In our experiments, this causal relation was kept constant across conditions, whereas the assignment of agent and patient roles was manipulated by either describing A as being capable of *sending* thoughts to B or B as being capable of *reading* the thoughts of A.

To test whether subjects have correct intuitions about the direction of causal dependency independent of the agent role assignment, we used a hypothetical intervention question in Experiment 1a. According to dependency theories, including causal Bayes nets, causal relations are asymmetric. Interventions in causes generate or prevent effects, whereas interventions in effects do not affect their causes (Pearl, 2000; Sloman, 2005; Waldmann & Hagmayer, 2005). Thus, a hypothetical manipulation of alien A's thoughts should have consequences for the thoughts of alien B but not vice versa if subjects invariantly view alien A as being involved in the cause event independent of whether alien A or alien B is assigned the agent role.

Experiment 1b used the same cover story as Experiment 1a but a different test question. In this study, we directly tested our hypothesis that subjects attribute responsibility differently in the sender and reader conditions.

3.1. Method

3.1.1. Participants and design

Two hundred and forty subjects (133 female; mean age 37.36 years) were recruited via an online database in the United Kingdom and compensated with an online voucher worth £ 0.50. Sixty-nine additional subjects were excluded from the analysis because they failed to pass a very simple logical transitivity task presented at the end of the experiment. We used this test to identify subjects who did not pay sufficient attention to the

task. Subjects were randomly assigned to one of four (Experiment 1a) or two (Experiment 1b) between-subjects conditions ($n = 40$).

3.1.2. Procedure and materials

The experiments were run as Web experiments using a tool for online studies (www.uni-park.de). In the instructions, subjects were told about two aliens, Gonz and Murks. It was pointed out that alien B, called Murks, often thinks of POR (indicated by a bubble containing POR) when alien A, called Gonz, thinks of POR (i.e., high causal strength) and that alien B occasionally thinks of POR for other reasons (i.e., a low but positive base rate). According to the instructions, POR refers to food in alien language. In the sender condition, subjects were told that alien A is capable of sending its POR thoughts to alien B. In the contrasting reader condition, subjects were told that alien B is able to read the POR thoughts of alien A. These instructions were explained in greater detail using the four possible combinations of aliens A and B thinking of POR or nothing. Up to this point, the instructions were identical in both experiments.

Then, in Experiment 1a (counterfactual intervention), subjects were asked to suppose they would have access to a mind-zapper machine that allows them to implant thoughts into an alien's mind (see also Steyvers et al., 2003). Furthermore, they were told to imagine a situation in which both aliens think of nothing. Two conditions were compared: In the *intervention-in-A condition*, subjects were asked to imagine implanting a POR thought in A's head and then to rate the probability of B thinking of POR (11-point scale from 0 to 10). In the contrasting *intervention-in-B condition,* subjects imagined implanting the POR thought in B's head and then rated the probability of A thinking of POR.

In Experiment 1b (responsibility attribution), subjects were asked to imagine a situation in which alien A thinks of POR, but alien B does not (hence, the thought transmission failed). Then, subjects were requested to indicate which of the two aliens, A or B, "is more responsible for this outcome" (forced choice between A and B).

3.2. Results and discussion

3.2.1. Experiment 1a: Counterfactual intervention

The results are shown in Fig. 5. When subjects hypothetically intervened in A, the ratings for the presence of POR thoughts in B were much higher than the ratings for POR thoughts in A when subjects imagined intervening in B, $F_{1,156} = 37.6$, $p < .001$, $\eta_p^2 = .19$. The agent role manipulation had no significant influence on the ratings, $F_{1,156} = 2.2$, $p = 0.14$, $\eta_p^2 = .01$.[3] As predicted, the interaction was not significant, $F_{1,156} = 1.2$, $p = .28$, $\eta_p^2 = .01$. Thus, regardless of the agent's involvement in the cause event (sender condition) or effect event (reader condition), the results indicate that subjects were aware of the causal dependency relation directed from cause (alien A) to effect (alien B) in both conditions. This finding provides evidence for our assumption that in both conditions subjects activated identical dependency intuitions despite the different placements of the agent.

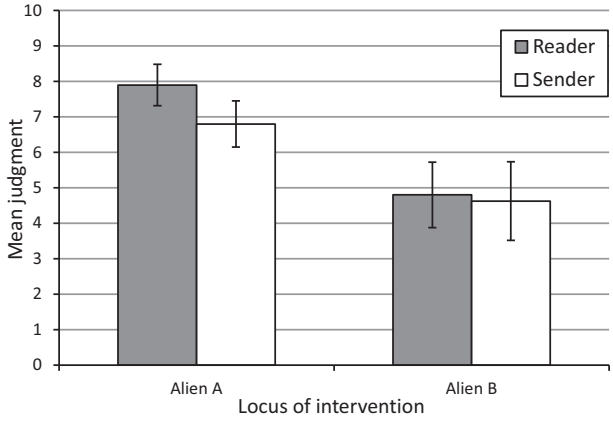

Fig. 5. Estimated probability (scale 0 to 10) of the presence of POR thoughts broken down by reader versus sender condition and locus of intervention (Experiment 1a).

3.2.2. Experiment 1b: Responsibility attribution

When asked to pick either alien A or alien B as responsible for a causal failure, 50% of the subjects (20 out of 40) chose alien A (i.e., cause event) in the sender condition, whereas in the contrasting reader condition alien A was only chosen by 17.5% of the subjects (7 of 40), $\chi^2_{1,80} = 9.45$, $p < .01$. This finding supports the hypothesis that agents attract error attributions. Interestingly, whereas in the reader condition choices were clearly asymmetric, the patient (i.e., alien B) was equally held responsible for failure in the sender condition. Apparently, subjects assumed in the sender condition that both sides contribute to the failure of transmission. Possibly, inferences in this condition were influenced by real-world analogies (e.g., radios) in which typically both sides of the transmission process can fail.

In sum, the most important finding for the following experiments is that the two conditions (i.e., sender vs. reader) activated the same dependency intuitions but clearly differed in terms of responsibility attribution. The following experiments tested the role of dispositional intuitions in more complex causal structures.

4. Experiment 2

For basic common-cause models with a single cause and several independent effects (e.g., Fig. 1), our model predicts strong Markov violations for scenarios in which the cause event is associated with the agent (e.g., sender condition) in contrast to scenarios in which it is associated with the patient (e.g., reader condition). To test this prediction, we presented subjects with instructions about four aliens, Gonz, Murks, Brxxx, and Zoohng, which either thought of POR (food in alien language) or of nothing. The POR thoughts of Gonz were described as probabilistically causing the POR thoughts of the three other aliens (i.e., common-cause model; see Fig. 1). In the test phase, subjects were requested

to rate the probability of a target effect alien thinking of POR given the thoughts of the cause alien, Gonz, and the other two effect aliens. Thus, as in Rehder and Burnett's (2005) paradigm, participants implicitly judged the conditional probability $P(E_3|C, E_1, E_2)$. As in Experiments 1a and 1b, we manipulated subjects' assumptions about the location of agents and patients: Gonz was either described as capable of sending its POR thoughts to the three effect aliens (sender condition) or the three effect aliens were described as capable of reading Gonz's POR thoughts (reader condition).

4.1. Method

4.1.1. Participants and design

Sixty students (44 female; mean age 23.9 years) of the University of Göttingen, Germany, participated in exchange for candy and were recruited on campus. Subjects were randomly assigned to one of two between-subjects conditions (sender vs. reader, $n = 30$).

4.1.2. Procedure and materials

The experiment consisted of two phases. In the *instruction phase*, subjects were presented with instructions about four aliens with one alien, Gonz, playing the role of a common cause of the thoughts of the three other aliens. It was pointed out that an effect alien often thinks of POR when Gonz thinks of POR (i.e., high causal strength) and that an effect alien occasionally thinks of POR for other reasons (i.e., low base rate of effects). In the sender condition, subjects were told that Gonz is able to send its POR thoughts to Murks, Brxxx, and Zoohng. In the contrasting reader condition, subjects were told that Murks, Brxxx, and Zoohng are able to read the POR thoughts of Gonz. These instructions were explained in greater detail showing the four possible cases involving two aliens thinking of POR or nothing. After reading the instructions, subjects were asked to summarize the task to make sure that they had understood it.

In the *test phase*, subjects were presented with six test cases in random order (for an example, see Fig. 6). In half of the cases, Gonz, the cause alien, thought of POR (indicated by POR being written in a bubble over its head); in the other half, it thought of nothing (indicated by an empty bubble). For each panel, subjects were asked to imagine 10 situations with the given configuration and then to judge in how many of these situations the target alien (indicated by a question mark above its head) would probably think of POR. This way, we obtained implicit probability assessments from each subject. We randomized across subjects which effect alien served as the target alien. Within subject, it was also manipulated how many of the two collateral effect aliens thought of POR or of nothing. All three possibilities—0, 1, and 2—were tested (yielding $2 \times 3 = 6$ within-subject conditions).

At the end of the experiment, subjects were asked to imagine a situation in which the cause is present, but two effects are absent (as depicted in Fig. 6). Then, we requested subjects to express their assessment of the relative degree of responsibility of the cause and effect aliens for this outcome by using a 5-point rating scale from −2 (*the cause alien*) to +2 (*the effect aliens*). We added this additional question to test whether the

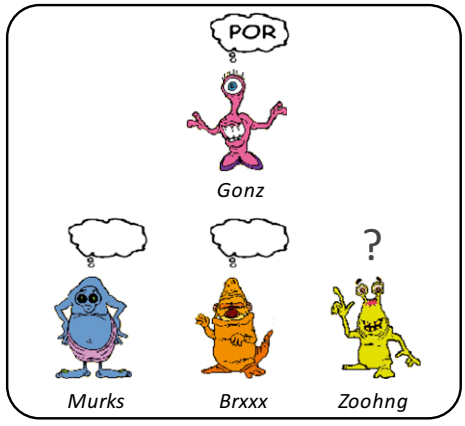

Fig. 6. An example of a test panel used in Experiment 2.

findings of Experiment 1b regarding differential error attributions generalize to more complex causal structures.

4.2. Results and discussion

The results of Experiments 2 are shown in Fig. 7 (see Fig. 4A for model predictions). In general, the ratings were higher when the cause was present (i.e., the cause alien thought of POR) than when the cause was absent (i.e., the alien did not think of POR), $F_{1,58} = 208.92$, $p < .001$, $\eta_p^2 = .78$. This is in line with the instructions mentioning high causal strength and low base rate of effects.

When the cause was present (the upper two lines in Fig. 7), a highly significant interaction was observed between the agency factor (reader vs. sender) and the number of collateral aliens thinking of POR, $F_{2,116} = 6.93$, $p < .01$, $\eta_p^2 = .10$; the three-way interaction across the states of the cause was also significant, $F_{2,116} = 3.19$, $p < .05$, $\eta_p^2 = .05$. This pattern reveals the predicted influence of the agent role assignment. The highly positive slope of the line in the sender condition (dashed line) indicates that subjects' ratings were strongly influenced by the collateral effects, $F_{2,58} = 13.06$, $p < .001$, $\eta_p^2 = .31$. Proportional to the number of collateral aliens failing to think of POR, the ratings for the target alien were lowered. This strong violation of the Markov condition is consistent with the prediction that the absence of collateral effects should be in part attributed to an error on the cause side, which should uniformly affect all effects (i.e., strong preventer F_C). In contrast, there was only a slight but not significant increase in the reader condition, $F_{2,58} = 1.26$, $p = .29$, $\eta_p^2 = .04$, indicating minimal violations of the Markov condition. This pattern shows that the failures of the collateral aliens to read the cause alien's thoughts were largely attributed to independent failures of their individual capacities and only to a small extent to the cause alien.

When the cause was absent (the lower two lines in Fig. 7), the ratings were uniformly low and did not differ across the sender and the reader conditions, $F_{2,116} < 1$. This is

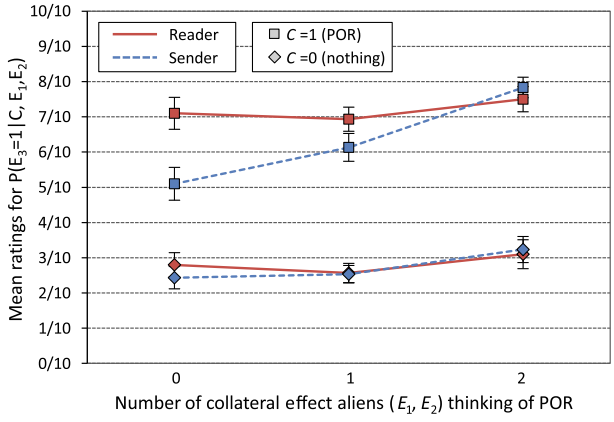

Fig. 7. Mean ratings (and standard errors) representing the estimates of the relative number of times the target alien thinks of POR in 10 hypothetical situations (Experiment 2).

consistent with the prediction of the model that in cases in which the cause is absent, agency role assignments should not influence judgments. However, we observed a very small, but reliable violation of independence, $F_{2,116} = 4.93$, $p < .01$, $\eta_p^2 = .08$. This finding is inconsistent with the model's prediction that effects should only be influenced by independent background factors when the cause is absent. However, because POR referred to food in alien language, it is possible that subjects inferred further (weak) common causes that influence food thoughts (e.g., external factors, such as the time of the day). In addition, the observation of multiple absent effects may have led to slight mutual adjustments of the base rate parameters. (In the modeling section, we have discussed an analogous possibility of mutual adjustments of causal strength parameters.) Such additional mechanisms could be integrated into our model, but were not the focus of our present research.

When asked to apportion responsibility to the aliens for the failure of the cause when two of its effects are absent (on a scale from −2 [*the cause alien*] to +2 [*the effect aliens*]), subjects attributed the failure slightly more to the cause alien in the sender condition ($M = -0.43$, $SD = 1.28$) and substantially more to the effect aliens in the reader condition ($M = 1.3$, $SD = 0.7$), yielding a significant difference between conditions, $t_{58} = 6.51$, $p < .001$, $d_{emp} = 1.68$. This pattern confirms that responsibility attributions are influenced by the location of agent and patient within the causal structure (see also Experiment 1b).

5. Experiment 3

In Experiment 2, we tested a causal situation in which the cause event varied between present and absent (i.e., 0/1 case). As predicted by our model, judgments were not affected by the location of agent and patient when the cause was absent. In the present

experiment, by contrast, we tested a situation in which the cause varied between two different present causally active values (i.e., A/B case; see also Rehder & Burnett, 2005, for similar materials). According to our model, both values should be considered causally effective and, hence, be subject to influences of the agency manipulation (sender vs. reader).

5.1. Method

5.1.1. Participants and design

Fifty-six students (29 female; mean age 25.1 years) from the University of Göttingen, Germany, participated in exchange for candy and were recruited on the campus. Subjects were randomly assigned to one of two between-subjects conditions (sender vs. reader, $n = 28$).

5.1.2. Procedure and material

As in Experiment 2, we presented subjects with instructions about four aliens that in this experiment had two possible different thoughts: They occasionally think of POR but usually think of TUS (indicated by a bubble containing POR or TUS; see Fig. 8; POR and TUS were counterbalanced across subjects). It was stated that the cause alien can transmit both thoughts to the three effect aliens (sender condition) or that the three effect aliens can read the thoughts of the cause alien (reader condition). Moreover, it was pointed out that the effect aliens frequently think of POR or TUS when the cause alien thinks of POR or TUS, respectively. Otherwise the instructions corresponded to those in Experiment 2.

In the test phase, we presented subjects with six test panels with all the nontarget aliens thinking of either POR or TUS (analogous to Experiment 2; see Fig. 8 for an example). The order of test panels was randomized. As in Experiment 2, judgments of the probability of POR thoughts of the target alien were collected for all constellations.

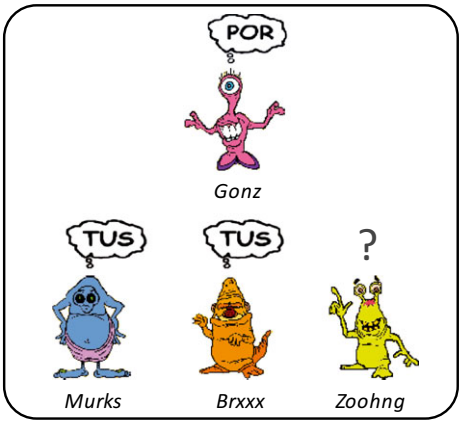

Fig. 8. An example of a test panel used in Experiment 3.

5.2. Results and discussion

The results of Experiment 3 are shown in Fig. 9 (see Fig. 4B for model predictions). As expected, the ratings for the target effect alien thinking of POR were higher when the cause alien thought of POR (and lower when the cause alien thought of TUS), $F_{1,54} = 146.05$, $p < .001$, $\eta_p^2 = .73$. As predicted by the model, subjects' judgments were influenced by the states of collateral effects for both states of the cause: In the case of C representing POR (the upper two lines in Fig. 9), the ratings substantially increased with the number of effect aliens thinking of POR, $F_{2,108} = 31.47$, $p < .001$, $\eta_p^2 = .37$. Replicating the findings of Experiment 2, this influence was stronger in the sender condition than in the reader condition, yielding a significant interaction, $F_{2,108} = 8.94$, $p < .001$, $\eta_p^2 = .14$. In the case of C representing TUS (the lower two lines in Fig. 9), the ratings also increased proportional to the number of effect aliens thinking of POR, $F_{2,108} = 20.25$, $p < .001$, $\eta_p^2 = .27$. Moreover, in contrast to what we observed in Experiment 2 in the cases in which the cause was absent, we also got a significant interaction when the cause alien did not think of POR (i.e., it thought of TUS), $F_{2,108} = 4.20$, $p < .05$, $\eta_p^2 = .07$. This descriptively weaker two-way interaction in the TUS case is predicted by the model as a consequence of the low base rate of POR (see Fig. 4B). These results provide further support for our model.

6. Experiment 4

In Experiment 4, we tested the predictions of our theory for causal chains. In this experiment, subjects were presented with a scenario in which four aliens form a chain in which each alien either sends thoughts to the next alien in the chain (sender condition) or

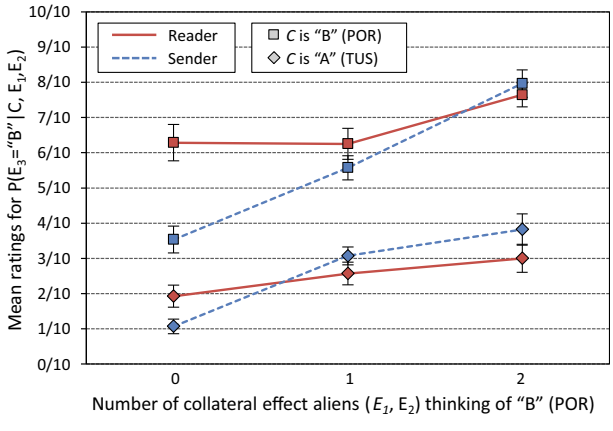

Fig. 9. Mean ratings (and standard errors) representing the estimates of the relative number of times the target alien thinks of POR in 10 fictitious situations (Experiment 3).

reads thoughts of the predecessor (reader condition). Unlike for common-cause structures, our model predicts for causal chains that the agency manipulation should not make a difference because now each cause alien carries its own preventive error node (as in the single-link case; see also Fig. 3D).

6.1. Method

6.1.1. Participants and design

Fifty students (22 female; mean age 24.3 years) from the University of Göttingen, Germany, participated in exchange for candy and were recruited on the campus. Subjects were randomly assigned to one of two between-subjects conditions (sender vs. reader, $n = 25$).

6.1.2. Procedure and material

As in Experiment 2, we presented subjects with instructions about four aliens, which most of the time think of nothing but sometimes think of POR (see Fig. 10, for an example of the spatial chain configuration). To instruct a causal chain, we either pointed out that each alien can transmit its POR thoughts to its right neighbor (sender condition) or that each alien can read the POR thoughts of its left neighbor (reader condition). Moreover, it was stated that aliens frequently think of POR when the left neighbor (i.e., its direct cause) also thinks of POR.

In the test phase, subjects were presented with six test panels with the nontarget aliens thinking of POR or nothing (for an example, see Fig. 10). The order of test panels was randomized. In all test questions, the target alien was the rightmost alien in the chain. As in Experiments 2 and 3, subjects were asked to estimate in how many of 10 situations the target alien would probably think of POR given the thoughts of the other aliens.

6.2. Results and discussion

The results of Experiment 4 are shown in Fig. 11. As in Experiments 2 and 3, the ratings for the target effect being present were higher in the presence of its direct cause (upper two lines in Fig. 11) than in its absence (lower two lines in Fig. 11), $F_{1,48} = 191.99$, $p < .001$, $\eta_p^2 = .80$. The prediction that different agency assignments

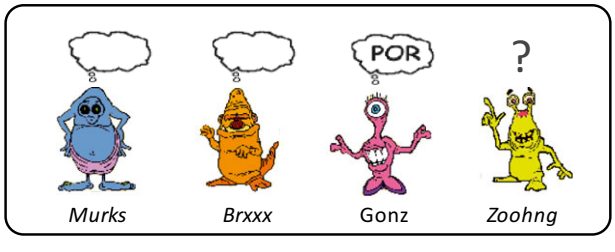

Fig. 10. An example of a test panel used in Experiment 4.

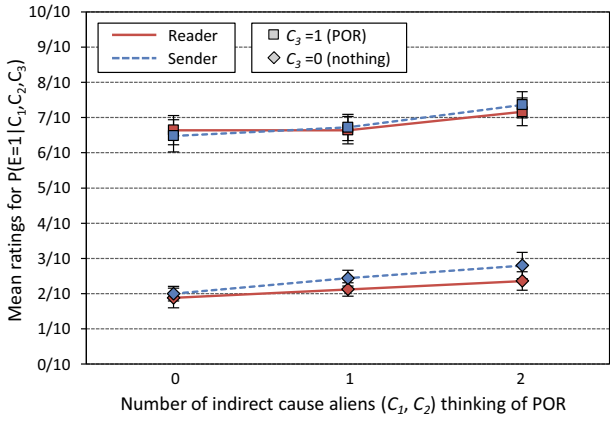

Fig. 11. Mean rating (and standard errors) of the number of times the target alien thinks of POR in 10 fictitious situations (Experiment 4).

within the chain (sender vs. reader) should not matter was clearly supported. In this experiment, no interaction between agency assignment and the states of the indirect causes was observed, neither in the presence of the direct cause of the target effect, $F_{2,96} < 1$, nor in its absence, $F_{2,96} < 1$. The three-way interaction was also not significant, $F_{2,96} < 1$. These results demonstrate that subjects adopted the instructed causal model and support the assumption inherent in our model that preventive nodes coding errors on the cause side (F_C) are attached to each cause separately.

However, qualitatively very weak violations of the Markov condition in both the direct cause's presence (the upper two lines in Fig. 11), $F_{2,96} = 11.77$, $p < .001$, $\eta_p^2 = .20$, and in its absence (the lower two lines in Fig. 11), $F_{2,96} = 6.47$, $p < .01$, $\eta_p^2 = .12$ were found. Markov violations are not prima facie predicted by our model for causal chains. However, Rehder and Burnett (2005) also found Markov violations in judgments about causal chains. One possible explanation for the weak trends violating the Markov condition in our experiment may be that some subjects may have doubted that chain variables fully screen off previous influences. As already discussed in the context of Experiment 2, it is also possible that subjects inferred further common causes that influence food thoughts (e.g., the time of the day) or mutually adjusted parameters (see Section 1.3.2, for a discussion of such a process). However, the most important prediction of our model that in causal chains the agency manipulation should not affect judgments was validated.

7. General discussion

According to dispositional theories (e.g., force dynamics), people classify causal participants into agents and patients. Causal agents are active entities endowed with the disposition to influence patients. Patients have the disposition to be influenced by agents, although they occasionally resist their influence. In contrast to dispositional theories,

dependency theories classify event variables into causes and effects. Although agents and causes were often confounded in previous research, we showed that these two categories are distinguishable. Separating causes and effects from agents and patients in our experiments allowed us to show that failures of causal transmission tend to be differentially attributed to agents and patients independent of whether they are located on the cause or effect side.

This finding constitutes a bridge between dispositional and dependency models. We have developed a causal Bayes net model that was augmented by separate error sources for causes and effects. Unlike in previous models in which error rates were collectively expressed in the causal strength parameters, the separation of different sources of error allowed us to model attributions of responsibility to agents and patients independently of the otherwise invariant causal dependency relations.

We tested this new hybrid model in several experiments using the size of Markov violations as an empirical indicator of differential assumptions about the sources of error. As predicted, we found larger Markov violations in common-cause models when the cause event involved the agent than when the agents were located on the effect side, with the exact pattern also influenced by the type of causal variables (Experiments 2 and 3). Another important finding, predicted by our model, was that agency assignments did not affect predictive inferences in causal chains (Experiment 4).

One interesting question for future research concerns the precise relation between mechanism knowledge and dispositional assumptions. In some simple cases or when subjects are domain experts, explicit mechanism knowledge may be fully available. In such cases, many subjects will use this knowledge to explain observed failures. However, more typical cases involve situations in which people have no or only partial mechanism knowledge (see Keil, 2012; Rozenblit & Keil, 2002). Most people know, for example, that aspirin has a disposition to remove headaches, and some may also have partial knowledge about the mechanisms. But many people will only know that some unknown chemicals inside the aspirin pill are responsible for the effect, or they will be able to name only some components of the mechanism (e.g., prostaglandins, ingestion, blood) without being able to fully trace the mechanism. We used alien mind reading as a test scenario to be able to study the role of *abstract* dispositional intuitions on causal inference independent of mechanism knowledge and expertise.

Obviously, it cannot be entirely ruled out that subjects activated intuitions based on analogies. All causal verbs are learned in the context of real-world cases that influence their semantic representations and may affect reasoning through analogies. People certainly have intuitions about sending and reading, but we suspect that few could elaborate exact mechanisms. As for our cover stories, the capacity of an alien to send or to read might just be represented as a dispositional placeholder for some intricate but unknown mechanism. One reviewer suggested that some subjects may split the cause variable into a component that represents the state of having the POR thought from a component representing the sending event. This hypothesis is in our view equivalent with our distinction between the presence of the cause (i.e., POR thought) and the activation of a potential inhibitor, the common preventive error node F_C, that can be

conceived of as an abstract placeholder for different potential factors underlying a failure of transmission on the cause side. Although our model shows that highly abstract intuitions about such factors suffice to explain inferences, it will certainly be interesting to further explore how different reasoners instantiate the placeholders with specific content.

A second interesting direction for future research concerns the relation between linguistic intuitions and causal reasoning (see Fausey & Boroditsky, 2010, for related research). To test our hypothesis that people differentially hold agents and patients accountable for causal failure, we have compared two different verbs, *send* and *read*. These two verbs allowed us to disentangle the agent and patient roles from the cause and effect events. We demonstrated that Markov violations almost disappeared in the reader condition. Given that *read* is a psych verb, one could postulate a general regularity between Markov violations and semantic verb categories. A clear mapping may, however, be elusive. Reading in the context of alien mind readers seems a more one-sided process than in the context of reading a book because people may have different intuitions about the role of the complementary participants. The alien whose thoughts are read might be seen as completely passive, whereas books are written with the intention to convey knowledge to readers. The assumption that the patient exerts no resistance during mind reading has certainly contributed to the unilateral attribution of error in the reader conditions in our experiments.

To achieve a general mapping between semantics and reasoning, it may be necessary to make finer grained semantic distinctions and to take into account additional context factors affecting our linguistic understanding. This hunch is confirmed by research on the implicit causality of causal verbs (see Rudolph & Försterling, 1997; for an overview). For example, although the psych verb *admire* typically suggests that the experienced stimulus carries the features causing admiration (i.e., B in "A admires B"), additional evidence suggesting that nobody else admires B shifts causal attributions toward A (Van-Kleeck, Hillger, & Brown, 1988).

Our research provides only an initial demonstration of how dispositional intuitions and causal reasoning within dependency models interact. We have focused on the role of agents and patients in error attribution to explain Markov violations in causal Bayes nets. More research, however, is needed to explore further interactions.

Acknowledgments

We wish to thank Marie-Theres Kater and Mira Holzer for assistance in data collection, and Marc Buehner, Noah Goodman, Tom Griffiths, York Hagmayer, Josh Tenenbaum, and two anonymous reviewers for helpful comments on the project. This work was supported by a travel grant from the Center for Statistics of the University of Göttingen and by the grants Wa 621/20 and Wa 621/22 from the Deutsche Forschungsgemeinschaft (DFG). The current project is part of the DFG priority program "New Frameworks of Rationality" (SPP 1516).

Notes

1. The qualitative pattern does not depend on the specific choice of parameters and is completely driven by the model's structure.
2. However, this mechanism could easily be implemented in our model.
3. The slight but nonreliable difference in ratings between agency conditions when subjects imagined intervening in A might express slightly different beliefs in causal strength for the instructed mechanisms (i.e., sending vs. reading).

References

Brown, R., & Fish, D. (1983). The psychological causality implicit in language. *Cognition*, *14*, 237–273.

Buchanan, D. W., Tenenbaum, J. B., & Sobel, D. M. (2010). Edge replacement and nonindepedence in causation. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the thirty-second annual conference of the cognitive science society* (pp. 919–924). Austin, TX: Cognitive Science Society.

Buehner, M. J., Cheng, P. W., & Clifford, D. (2003). From covariation to causation: A test of the assumption of causal power. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 1119–1140.

Cartwright, N. (2007). *Hunting causes and using them: Approaches in philosophy and economics*. Cambridge, UK: Cambridge University Press.

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

Cheng, P. W., & Novick, L. R. (1992). Covariation in natural causal induction. *Psychological Review*, *99*, 365–382.

Cushman, F. (2008). Crime and punishment: Distinguishing the roles of causal and intentional analyses in moral judgment. *Cognition*, *108*, 353–380.

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, *67*, 547–619.

Fausey, C. M., & Boroditsky, L. (2010). Subtle linguistic cues influence perceived blame and financial liability. *Psychonomic Bulletin & Review*, *17*, 644–650.

Gerstenberg, T., & Lagnado, D. A. (2010). Spreading the blame: The allocation of responsibility amongst multiple agents. *Cognition*, *115*, 166–171.

Goldvarg, E., & Johnson-Laird, P. N. (2001). Naïve causality: A mental model theory of causal meaning and reasoning. *Cognitive Science*, *25*, 565–610.

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.

Hausman, D. M., & Woodward, J. (1999). Independence, invariance and the causal Markov condition. *British Journal for the Philosophy of Science*, *50*, 521–583.

Hausman, D. M., & Woodward, J. (2004). Modularity and the causal Markov condition: A restatement. *British Journal for the Philosophy of Science*, *55*, 147–161.

Holyoak, K. J., & Cheng, P. W. (2011). Causal learning and inference as a rational process: The new synthesis. *Annual Review of Psychology*, *62*, 135–163.

Hume, D. (1748/1977). *An enquiry concerning human understanding*. Indianapolis, IN: Hackett Publishing Company.

Keil, F. C. (2012). Running on empty? How folk science gets by with less. *Current Directions in Psychological Science*, *21*, 329–334.

Kistler, M., & Gnassounou, B. (Eds.) (2007). *Dispositions and causal powers*. Aldershot, UK: Ashgate.

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

Landau, I. (2010). *The locative syntax of experiencers*. Cambridge, MA: MIT Press.

Levin, B., & Rappaport Hovav, M. (2005). *Argument realization*. Cambridge, MA: Cambridge University Press.

Lewis, D. (1973). Causation. *Journal of Philosophy*, *70*, 556–567.

Lien, Y., & Cheng, P. W. (2000). Distinguishing genuine from spurious causes: A coherence hypothesis. *Cognitive Psychology*, *40*, 87–137.

López, F. J., & Shanks, D. R. (2008). Models of animal learning and their relations to human learning. In R. Sun (Ed.), *Cambridge handbook of computational psychology* (pp. 589–611). Cambridge, UK: Cambridge 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.

Mayrhofer, R., Goodman, N. D., Waldmann, M. R., & Tenenbaum, J. B. (2008). Structured correlation from the causal background. In B. C. Love, K. McRae, & V. M. Sloutsky (Eds.), *Proceedings of the thirtieth annual conference of the cognitive science society* (pp. 303–308). Austin, TX: Cognitive Science Society.

Mayrhofer, R., Hagmayer, Y., & Waldmann, M. R. (2010). Agents and causes: A Bayesian error attribution model of causal reasoning. In S. Ohlsson & R. Catrambone (Eds.), *Proceedings of the thirty-second annual conference of the cognitive science society* (pp. 925–930). Austin, TX: Cognitive Science Society.

Mayrhofer, R., & Rothe, A. (2012). Causal status meets coherence: The explanatory role of causal models in categorization. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the thirty-fourth annual conference of the cognitive science society* (pp. 743–748). Austin, TX: Cognitive Science Society.

Mayrhofer, R., & Waldmann, M. R. (2011). Heuristics in covariation-based induction of causal models: Sufficiency and necessity priors. In C. H. Carlson, & T. Shipley (Eds.), *Proceedings of the thirty-third annual conference of the cognitive science society* (pp. 3110–3115). Austin, TX: Cognitive Science Society.

Mayrhofer, R., & Waldmann, M. R. (2013). Agency intuitions in physical interactions. In M. Knauff, M. Pauen, N. Sebanz, & I. Wachsmuth (Eds.), *Proceedings of the 35th annual conference of the cognitive science society* (pp. 996–1001). Austin, TX: Cognitive Science Society.

Meder, B., Mayrhofer, R., & Waldmann, M. R. (in press). Structure induction in diagnostic causal reasoning. *Psychological Review*.

Michotte, A. E. (1963). *The perception of causality*. New York: Basic Books.

Mumford, S. (2003). *Dispositions*. Oxford, UK: Oxford University Press.

Mumford, S., & Anjum, R. L. (2011). *Getting causes from powers*. New York: Oxford University Press.

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

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

Perales, J. C., Catena, A., & Maldonado, A. (2004). Inferring non-observed correlations from causal scenarios: The role of causal knowledge. *Learning and Motivation*, *35*, 115–135.

Pinker, S. (1991). Rules of language. *Science*, *253*(5019), 530–535.

Pinker, S. (2007). *The stuff of thought: Language as a window into human nature*. New York: Viking.

Rehder, B. (2003a). Categorization as causal reasoning. *Cognitive Science*, *27*, 709–748.

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

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

Riemer, N. (2010). *Introducing semantics*. Cambridge, UK: Cambridge University Press.

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

Rozenblit, L., & Keil, F. C. (2002). The misunderstood limits of folk science: An illusion of explanatory depth. *Cognitive Science*, *26*, 521–562.

Rudolph, U., & Försterling, F. (1997). The psychological causality implicit in verbs: A review. *Psychological Bulletin*, *121*, 192–218.

Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton, NJ: Princeton University Press.

Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in describing persons: Social cognition and language. *Journal of Personality and Social Psychology*, *54*, 558–568.

Semin, G. R., & Fiedler, K. (1991). The linguistic category model, its bases, applications and range. *European Review of Social Psychology*, *2*, 1–30.

Shanks, D. R., & Darby, R. J. (1998). Feature- and rule-based generalization in human associative learning. *Journal of Experimental Psychology: Animal Behavior Processes*, *24*, 405–415.

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

Spirtes, P., Glymour, C., & Scheines, P. (1993). *Causation, prediction, and search*. New York: Springer.

Spohn, W. (2012). *Ranking theory: A tool for epistemology*. Oxford, England: Oxford University Press.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E.-J., & Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive Science*, *27*, 453–489.

Talmy, L. (1988). Force dynamics in language and cognition. *Cognitive Science*, *12*, 49–100.

VanKleeck, M. H., Hillger, L. A., & Brown, R. (1988). Pitting verbal schemas against information variables in attribution. *Social Cognition*, *6*, 89–106.

Waldmann, M. R. (1996). Knowledge-based causal induction. In D. R. Shanks, K. J. Holyoak, & D. L. Medin (Eds.), *The psychology of learning and motivation* (Vol. 34, pp. 47–88). San Diego, CA: Academic Press.

Waldmann, M. R., & Hagmayer, Y. (2005). Seeing versus doing: Two modes of accessing causal knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 216–227.

Waldmann, M. R., & Hagmayer, Y. (2013). Causal reasoning. In D. Reisberg (Ed.), *Oxford handbook of cognitive psychology* (pp. 733–752). New York: Oxford University Press.

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General*, *121*, 222–236.

Waldmann, M. R., Holyoak, K. J., & Fratianne, A. (1995). Causal models and the acquisition of category structure. *Journal of Experimental Psychology: General*, *124*, 181–206.

Waldmann, M. R., & Martignon, L. (1998). A Bayesian network model of causal learning. In M. A. Gernsbacher, & S. J. Derry (Eds.), *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society* (pp. 1102–1107). Mahwah, NJ: Erlbaum.

Walsh, C. R., & Sloman, S. A. (2007). Updating beliefs with causal models: Violations of screening off. In M. A. Gluck, J. R. Anderson, & S. M. Kosslyn (Eds.), *Memory and mind: A Festschrift for Gordon H. Bower* (pp. 345–357). New York: Laurence Erlbaum.

White, P. A. (2006). The causal asymmetry. *Psychological Review*, *113*, 132–147.

White, P. A. (2007). Impressions of force in visual perception of collision events: A test of the causal asymmetry hypothesis. *Psychonomic Bulletin & Review*, *14*, 647–652.

White, P. A. (2009). Perception of forces exerted by objects in collision events. *Psychological Review*, *116*, 580–601.

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

Wolff, P. (2012). Representing verbs with force vectors. *Theoretical Linguistics*, *38*, 237–248.

Wolff, P., Barbey, A. K., & Hausknecht, M. (2010). For want of a nail: How absences cause events. *Journal of Experimental Psychology: General*, *139*, 191–221.

Wolff, P., & Song, G. (2003). Models of causation and the semantics of causal verbs. *Cognitive Psychology*, *47*, 276–332.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. New York: Oxford University Press.

Woolfolk, R. L., Doris, J. M., & Darley, J. M. (2006). Identification, situational constraint, and social cognition: Studies in the attribution of moral responsibility. *Cognition*, *100*, 283–301.