A cognitively plausible algorithm for causal inference

Gordon Briggs and Sangeet Khemlani
{gordon.briggs, sangeet.khemlani}@nrl.navy.mil
Navy Center for Applied Research in Artificial Intelligence
US Naval Research Laboratory, Washington, DC 20375 USA

Abstract

People without any advanced training can make deductions about abstract causal relations. For instance, suppose you learn that *habituation causes seriation*, and that *seriation prevents methylation*. The vast majority of reasoners infer that *habituation prevents methylation*. Cognitive scientists disagree on the mechanisms that underlie causal reasoning, but many argue that people can mentally simulate causal interactions. We describe a novel algorithm that makes domain-general causal inferences. The algorithm constructs small-scale iconic simulations of causal relations, and so it implements the "model" theory of causal reasoning (Goldvarg & Johnson-Laird, 2001; Johnson-Laird & Khemlani, 2017). It distinguishes between three different causal relations: causes, enabling conditions, and preventions. And, it can draw inferences about both orthodox relations (*habituation prevents methylation)* and omissive causes (*the failure to habituate prevents methylation*). To test the algorithm, we subjected participants to a large battery of causal reasoning problems and compared their performance to what the algorithm predicted. We found a close match between human causal reasoning and the patterns predicted by the algorithm.

Keywords: causation; mental models; reasoning; simulation

Introduction

People routinely make inferences about complex causal matters. For instance, consider the following description about a particular farm:

1. Flourishing weeds will cause a lack of nutrients.
   A lack of nutrients will prevent the vegetables from growing.
   The lack of vegetables will enable an early harvest.

What is the relation between the growth of weeds and an early harvest? Reasoners needn't have a background in botany to infer a possible causal relation between the two events, such as in (2):

2. Flourishing weeds will cause an early harvest.

People's inferences are systematic, and at least some errors are obvious, i.e., anyone who infers (3) from the information in the description above is grossly mistaken:

3. Flourishing weeds will prevent an early harvest.

How do people infer causal relations between events? Sometimes, perceptual cues may drive people to infer a causal connection between one event and another: if you observe that when a man flips a switch, a particular light goes off, it seems reasonable to infer a causal relation between the switch and the light. Indeed, the temporal contiguity of two events can be sufficient to imply causality (e.g., Lagnado & Sloman, 2006; Rottman & Keil, 2012). But the preceding farming example demonstrates that people can infer causal relations from descriptions, not just observations, and that they can do so in the absence of any explicit temporal information.

How do people make causal inferences? A popular approach in artificial intelligence simulates human causal reasoning using causal Bayes nets and a calculus developed by Pearl (2009). It allows precise calculations of conditional probabilities, e.g., the probability of an early harvest given flourishing weeds, P(*early harvest* | *flourishing weeds*), provided that relevant causal relations are translated into the notation of a graphical network. While the approach can distinguish between causes and mere associations, Pearl's calculus cannot explain how reasoners infer *novel* causal relations where none had existed before, i.e., it cannot explain how people infer (2) from (1).

Cognitive scientists disagree on the mechanisms and representations that underlie causal reasoning (Ahn & Bailenson, 1996; Cheng, 1997; Sloman, 2005; White, 2014; Wolff & Barbey, 2015). Mental simulation is central to many psychological accounts of the process: theorists agree that people construct small-scale simulations to predict outcomes (Kahneman & Tversky, 1982), to understand mechanistic relations (Hegarty, 2004), to comprehend physical scenes (Battaglia, Hamrick, & Tenenbaum, 2013), to resolve inconsistent and contradictory information (Khemlani & Johnson-Laird, 2011), to deduce the consequences of sequences of events (Khemlani, Mackiewicz, Bucciarelli, & Johnson-Laird, 2013), and to make counterfactual inferences (Byrne, 2005; Galinsky & Moskowitz, 2000).

Recent approaches to modeling causal reasoning in AI and cognitive science face two overarching challenges: first, people distinguish between causal relations such as *cause*, *enable*, and *prevent*. They understand, for instance, that (4a) and (4b) mean different things:

4a. A lack of vegetables will *cause* an early harvest.
 b. A lack of vegetables will *enable* an early harvest.

Graphical networks have difficulty capturing the difference between the two relations. Various psychological theories have invoked the transmission of causal forces (Wolff, 2007), causal model structures (Sloman et al., 2009), and mental simulations of possibilities (Goldvarg & Johnson-Laird, 2001) to explain what different causal relations mean (for a review, see Khemlani, Barbey, & Johnson-Laird, 2014). But there exists no robust computational model that predicts what causal relations people generate from descriptions such as (1) above.

Second, most theories of causal reasoning cannot explain reasoning about *omissive* causal relations, such as in (5):

5. A lack of nutrients will cause the vegetables to die.

The assertion is distinct from (2) above because it describes how the absence of an element can bring about some outcome. Philosophers, psychologists, and computer scientists have so much difficulty coping with omissive causation that some philosophers deny it as a meaningful concept (e.g., Beebee, 2004; Dowe, 2001; Hall, 2004). In recent years, psychologists advanced theories to account for omissive causation: some theorists treat omissive causes as an arrangement of causal forces (Wolff, Barbey, & Hausknecht, 2010) or as a set of counterfactual contrasts (Stephan, Willemsen, & Gerstenberg, 2017). But, counterfactuals cannot explain how people reason about future causal relations, such as in (1) above, because the counterfactuals are retrospective by definition. And, forces do not explain why reasoners appear to distinguish omissive causes from omissive enabling conditions and omissive preventions (see, e.g., Khemlani, Wasylyshyn, Briggs, & Bello, 2018).

Hence, students of causal reasoning remain bereft of a feasible, adequate process model of how humans infer causal relations. Our goal in the present article is to specify such an algorithm and to describe its computational implementation. The algorithm is based on the notion that people build iconic simulations of possibilities when they reason, and that they mentally scan those possibilities to infer novel conclusions. Since the goal of the algorithm is to account for human intuitions, we describe an experiment used to benchmark the algorithm, and we show how its implementation matches the performance of human reasoners. We also describe a set of simulations used to validate the parameters in the implementation. We conclude by evaluating the results in the context of contemporary accounts of causal reasoning.

Mental models and causal reasoning

The algorithm for causal inference we present in this paper is based on the tenets of mental model theory – the "model" theory for short. The model theory argues that reasoning depends on the mental simulation of sets of possibilities. The theory is based on three fundamental principles:

- **Mental models represent possibilities.** When people reason about relations, causal or otherwise, they construct one or more possibilities – situations describing finite alternatives – consistent with those relations (Johnson-Laird, 2006; Khemlani, Byrne, & Johnson-Laird, 2018).
- **Mental models are iconic.** The structure of a mental model mirrors the structure of what it represents as far as possible (Peirce, 1931-1958, Vol. 4). An iconic simulation of a causal relation, e.g., *A causes B,* concerns sets of events, *A* and *B*, in a temporal order. Models can also include abstract symbols, e.g., the symbol for negation (Khemlani, Orenes, & Johnson-Laird, 2012) and they can represent sequences of events as they unfold in time (Khemlani et al., 2013).
- **Intuitions depend on one model; deliberations depend on multiple models.** Human reasoning is based on two interacting sets of processes: people form rapid, intuitive inferences by constructing and scanning a single "mental" model, but those intuitive inferences lead individuals to make errors (Khemlani & Johnson-Laird, 2017). Mistakes can be corrected by deliberation, which requires reasoners to consider multiple models by searching for counterexamples to intuitive conclusions (Khemlani & Johnson-Laird, 2013; Khemlani et al., 2015).

The model theory explains why people distinguish between *causes, enables*, and *prevents:* each relation refers to a distinct set of possibilities (Goldvarg & Johnson-Laird, 2001), known as *fully explicit models*. Table 2 shows the fully explicit models for the three relations. For instance, a causal assertion such as (2) above refers to a conjunction of three separate models of possibilities, depicted in this schematic diagram:

```
  weeds       early-harvest
¬ weeds       early-harvest
¬ weeds     ¬ early-harvest
```

Each row in the diagram represents a distinct temporally ordered possibility, e.g., the first row represents the possibility in which weeds flourish and then an early harvest occurs. Any possibility missing from the diagram is inconsistent with (2): hence, the situation in which weeds occur and an early harvest does not is incompatible with (2), and so too is any possibility in which an early harvest occurs before the weeds flourish. In contrast, an enabling condition, such as in (6):

6. Flourishing weeds will enable aphids to thrive.

refers to a different conjunction of possibilities:

```
  weeds       aphids
  weeds     ¬ aphids
¬ weeds     ¬ aphids
```

Unlike causes, enabling conditions permit the situation in which the antecedent occurs but the consequent doesn't, e.g., (6) allows for the possibility in which weeds occur but aphids don't thrive. Typically, enabling conditions rule out the possibility in which aphids thrive in the absence of weeds. As Goldvarg and Johnson-Laird (2001) showed, reasoners list these possibilities for assertions such as (2) and (6). Reasoning about causal relations, however, requires significantly more processing than interpreting causal statements, and so when people have to reason, they often do not consider the full list of possibilities – instead, they draw conclusions from just a single possibility, referred to as the *mental model*. The mental models for causes and enabling conditions are bolded in the diagrams above. They're identical, and as a result, individuals often fail to distinguish enabling from causing when they reason (see Experiment 5 in Goldvarg & Johnson-Laird, 2001). Preventions are akin to causes with a negated consequent (see Table 1).

A recent development of the model theory shows that it can explain omissions: the theory treats them as negated antecedents (Khemlani et al., 2018). Hence, the fully explicit models for (5) above are:

```
¬ nutrients     dying-vegetables
  nutrients     dying-vegetables
  nutrients   ¬ dying-vegetables
```

Analogous changes explain omissive enabling conditions and omissive preventions (Table 1). The theory accordingly uses a unified representation for both omissions and orthodox causes.

The model theory explains how people represent causal relations, and various empirical assessments validate the theory's central predictions (Johnson-Laird & Khemlani, 2017). We turn next to describe a novel algorithm and its computational implementation, and we show how to compute inferences from models of possibilities.

Computing with mental models

The algorithm used to infer causal relations relies on three separate subroutines, each of which depends on the representational conventions described in the previous section. First, the algorithm needs to build integrated models of multiple causal relations, e.g., it needs to combine the three sentences in (1) above into a set of models. Second, since reasoners are unlikely to construct models deterministically, the algorithm needs to specify a stochastic system that can mimic the distribution of possible interpretations that humans tend to make. Third, the algorithm needs to explain how people scan models to generate novel relations. We review each subroutine in turn.

Building integrated models

To construct an integrated model from a set of premises, the algorithm adopts a mechanism developed for previous model-theoretic computational implementations (see, e.g., Johnson-Laird & Byrne, 1991): the algorithm takes the Cartesian product of two models with the proviso that a model of an event cannot be combined with its negation. An example will illustrate the process. Consider the premises in (7), both of which concern omissions:

7. A lack of sunlight will prevent the vegetables from growing.
   The lack of vegetables will enable an early harvest.

The mental model of the first premise is:

```
¬ sunlight   ¬ vegetables
```

and the model of the second is:

```
             ¬ vegetables     early-harvest
```

So, a Cartesian product of the two models identifies that the middle event is shared in both models, and it combines the two to create an integrated model:

```
¬ sunlight   ¬ vegetables     early-harvest
```

Suppose instead that people build fully explicit models of the premises, not mental models. The fully explicit model of the first premise is:

```
¬ sunlight   ¬ vegetables
  sunlight   ¬ vegetables
  sunlight     vegetables
```

and the fully explicit model of the second premise is:

```
¬ vegetables     early-harvest
¬ vegetables   ¬ early-harvest
  vegetables   ¬ early-harvest
```

A procedure implementing the Cartesian product starts by combining the first model of the first premise with the three models of the second premise to yield:

```
¬ sunlight   ¬ vegetables     early-harvest
¬ sunlight   ¬ vegetables   ¬ early-harvest
```

The last model of the second premise is a situation in which vegetables grow, and so it cannot be combined with the first model of the first premise. The same procedure applies to the second and third models of the first premise, and so the full Cartesian product of the two sets of fully explicit models is:

```
8. ¬ sunlight   ¬ vegetables     early-harvest
   ¬ sunlight   ¬ vegetables   ¬ early-harvest
     sunlight   ¬ vegetables     early-harvest
     sunlight   ¬ vegetables   ¬ early-harvest
     sunlight     vegetables   ¬ early-harvest
```

Reasoners are likely to vary in their tendency to interpret causal assertions using mental models or fully explicit models, and so the algorithm implementing the theory uses a stochastic parameter to govern the process: the ε-parameter determines the probability that the algorithm will construct only the mental model or whether it will construct fully explicit models (see, e.g., Johnson-Laird, Khemlani, & Goodwin, 2015; Khemlani & Johnson-Laird, 2013, 2016; Khemlani et al., 2015; for applications of this methodology to quantificational reasoning). The parameter accordingly varies from 0.0 to 1.0 such that when $\varepsilon = 0.0$, the algorithm always produces mental models, and when $\varepsilon = 1.0$, the algorithm produces fully explicit models. Hence, the ε parameter varies the contents of the models.

Varying the size of models

Another parameter, the λ-parameter, controls the number of possibilities that the algorithm yields as it constructs an integrated model. It therefore controls size of the models. This parameter corresponds to the λ-parameter of a Poisson distribution. Consider how the parameter might apply to interpreting the premises in (7). On any given run of the the algorithm, the size of a set of models is governed by $n_{\text{Premise 1}} + n_{\text{Premise 2}}$, both of which are established by two samples drawn from a Poisson distribution of parameter λ. Once the two ns are determined, possibilities are sampled from the fully explicit models and their Cartesian product is taken to yield an integrated mental model. Hence, if $n_{\text{Premise 1}} = 2$, the algorithm would sample 2 separate possibilities from the

	Conjunctions of possibilities yielding different causal relations					
	A causes B	*A enables B*	*A prevents B*	*Not A causes B*	*Not A enables B*	*Not A prevents B*
Fully explicit models	A B	A B	A ¬B	¬A B	¬A B	¬A ¬B
	¬A B	A ¬B	¬A ¬B	A B	¬A ¬B	A ¬B
	¬A ¬B	¬A ¬B	¬A B	A ¬B	A ¬B	A B
Mental model	A B	A B	A ¬B	¬A B	¬A B	¬A ¬B

Table 1. The possibilities consistent with various causal relations in the model theory. Reasoners distinguish between the meanings of relations based on the distinct sets of possibilities – the *fully explicit models* – to which they refer. But, when they make inferences, people often consider just one of the possibilities consistent with the meaning of a relation – the *mental model*. Background knowledge can block the construction of certain models, e.g., *alcohol causes inebriation* is true, and since only alcohol causes inebriation, people should not consider the situation in which inebriation occurs in the absence of alcohol, i.e., the ¬A B model in the first column. A more thorough discussion of strong and weak interpretations is provided in Johnson-Laird and Khemlani (2017).

3 consistent with *not-A prevents B*, which corresponds to the first premise of (7). The same procedure would be used for the second premise. Their Cartesian product would be taken, and since the product concerns sets of fewer possibilities, the resulting integrated model would be a subset of the models in (8) above, e.g.,

```
9. ¬ sunlight   ¬ vegetables     early-harvest
   ¬ sunlight   ¬ vegetables   ¬ early-harvest
     sunlight     vegetables   ¬ early-harvest
```

The algorithm provides two distinct methods of sampling from the possibilities to which the relations refer: the first method samples *n* separate possibilities uniformly; the second samples the possibilities in the order specified by Table 1. Previous empirical results suggest that reasoners list certain possibilities more frequently than others in a manner predicted by the model theory (Bello, Wasylyshyn, Briggs, & Khemlani, 2017). A simulation analysis presented below tests whether random sampling or preferential sampling produces a better to human data.

Generating causal inferences

To generate causal inferences from, e.g., an integrated model such as (9) above, the algorithm reduces the integrated model to a model of its end terms, discarding redundant models where relevant. The reduction process for (9) yields the model in (10) below:

```
10.¬ sunlight                    early-harvest
   ¬ sunlight                  ¬ early-harvest
     sunlight                  ¬ early-harvest
```

The algorithm attempts to match this reduced set of possibilities with all combinations of possibilities in Table 1. If one or more matches can be found in Table 1, the algorithm can form a response by choosing randomly from the corresponding matching relations. In the case of (10), matching relations include*: sunlight prevents an early harvest* and *a lack of sunlight enables an early harvest*.

A more sophisticated response heuristic integrated into the algorithm assesses the first premise of the problem to check whether the antecedent it describes concerns omissive or orthodox causation. For (7), the antecedent – "a lack of sunlight" – concerns omission, the only candidate response is: *a lack of sunlight enables an early harvest.*

To assess whether the algorithm we describe matches human causal reasoning responses, we collected data from participants and compared their responses to those generated by the algorithm.

Experiment and simulations

We conducted an experiment to test the algorithm specified in the previous section. The experiment replicated a design developed by Wolff and Barbey (2015, Experiment 3), in which the authors provided participants with 32 causal reasoning problems of the following form:

X prevents Y.
Y prevents Z.
What, if anything, follows?

In their original study, participants carried out a multiple-choice task in which they selected which responses followed of necessity from 10 possible options. Multiple-choice tasks are limited in their ecological validity – the task encourages participants to select multiple responses, and the order in which they select those responses is subject to carry-over effects. To address the limitation, we replicated their design but used a fill-in-the-blank task to test participants' natural responses to causal reasoning problems. Participants in our study registered their responses by using a series of dropdown menus to formulate a conclusion that relates *X* and *Z*:

[X/¬X] [causes/enables/prevents] [Z/¬Z]

Participants provided one response to each problem.

Method

Participants. 50 participants were recruited through Amazon Mechanical Turk (28 male, mean age = 34.6). 15 participants reported some formal logic or advanced training in mathematics, and all but 1 of the participants were native English speakers.

Design, procedure, and materials. Each participant was presented with 32 two-premise causal inference problems taken from Wolff and Barbey (2015). The causes and effects in each premise were populated from a set of fictional conditions (e.g., "valmork temperaments", "kandersa moods"). Orthodox and omissive antecedents were created using the phrases “having” and “not having,” respectively, and so some participants received the following problem:

Having valmork temperaments prevents kandersa disease.
Having kandersa disease prevents rempust fever.

The order in which the participants carried out the 32 problems was randomized, as was the assignment of the contents of the premises. Data, materials, experimental code, and computational modeling code are available at https://osf.io/5yqfx.

Results and simulations

Figure 1 (top panel) shows the data from the experiment. As the figure shows, different problems yielded markedly different patterns of response, e.g., participants generated the response “Not X causes Z” for only one of the 32 problems. For brevity, we omit further analyses of the experimental data in favor of using the dataset to benchmark a series of simulation analyses.

Four separate versions of the algorithm were implemented. The four versions reflected the two strategies for model constructed described above (*random sampling* or *preference sampling*) and the two sorts of response policy (*random selection* or *heuristic selection*). A separate parameter search was conducted for each of the four versions of the algorithm.

Sampling method	Response selection	Best fitting ε value	Best fitting λ value	Goodness of fit (*r*)
Random	Random	0.8	1.0	.65
Random	Heuristic	0.8	1.3	.71
Preferential	Random	0.9	0.9	.71
Preferential	**Heuristic**	**1.0**	**0.8**	**.75**

Table 2. The model-fitting results of simulation analyses conducted for each of the four versions of a model-based causal reasoning algorithm. The version of the algorithm that used preferential sampling and heuristic response generation yielded the best fit to the data.

For each parameter search, the parameters ε and λ varied in 0.1 increments such that the ε ranged from 0.0 to 1.0 and the λ parameter ranged from 0.0 to 3.0, which produced 300 separate parameter configurations. For each parameter configuration, the algorithm was run 100 times on each of the 32 separate causal reasoning problems.

Table 2 compares the overall results of each of the four versions of the algorithm. The table shows that the version of the algorithm that used preferential sampling to construct integrated models as well as a heuristic response strategy performed better than the other three versions of the algorithm. Figure 1 (bottom panel) shows the data generated by the best fitting simulation amongst the four versions of the algorithm.

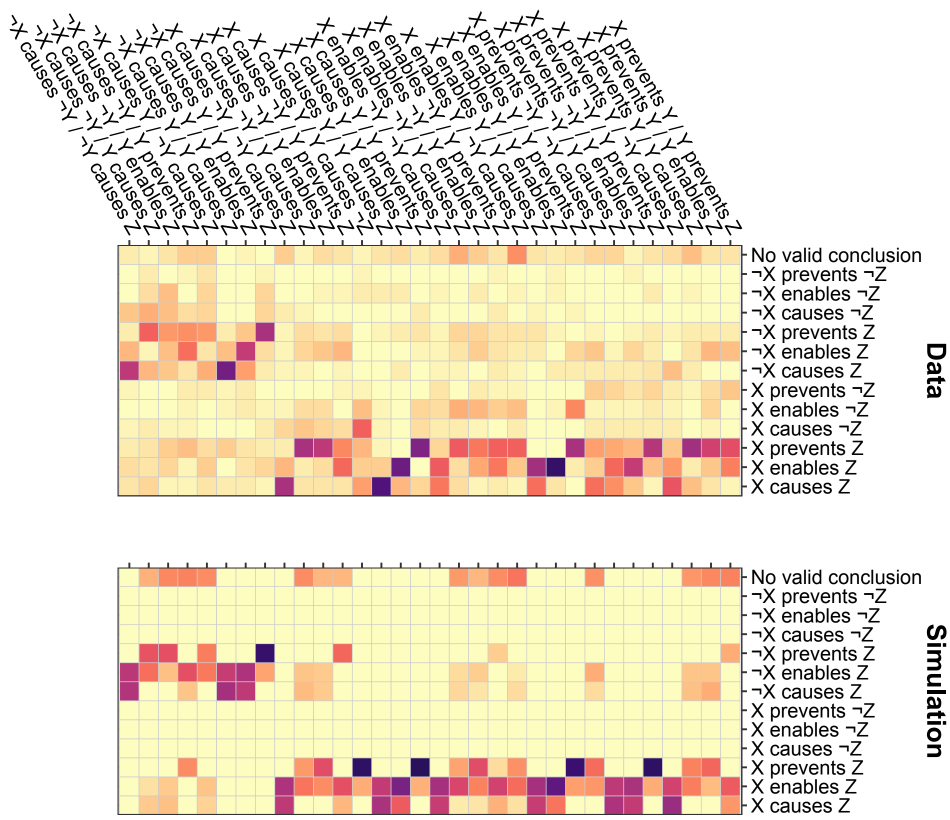

Figure 1. The proportions of participants’ responses to 32 different causal reasoning problems (top panel) and from the best fitting simulation from the algorithm that computes causal inferences (bottom panel). The color in each cell indicates the proportion of corresponding conclusions such that the darker the cell, the higher the proportion. Hence, nearly 100% of participants responded “X enables Z” when responding to the problem: “X enables Y / Y enables Z”. The version of the algorithm that yielded the best fit implemented a preferential sampling and a heuristic response selection policy.

To assess the necessity of the algorithm's two parameters, we carried out parameter lesioning tests for the version of the algorithm that used preferential sampling and heuristic response selection. Specifically, we ran the algorithm in two lesioned conditions: one in which ε was set to 0, while λ was permitted to vary, and another in which λ was set to 4.0 while ε was permitted to vary. If either condition performed as well as the optimal fit, then it suggests that one of the parameters was redundant. But, neither lesioned condition produced a better fit to the data: the best fitting simulation when λ was permitted to vary yielded a lower goodness-of-fit (r = .64) and likewise for the best fitting simulation when ε was permitted to vary (r = .44). We conclude that the algorithm that incorporated preferential sampling and heuristic response generation produced the closest match to participants' inferences (r = .75).

General discussion

We introduced a novel algorithm for computing causal inferences from sets of causal premises. The algorithm mimics human inference because it is based on a cognitive theory of reasoning, the model theory (Khemlani et al., 2014). It generates causal conclusions by following three procedures: first, the system stochastically constructs mental models from the meanings of causal relations. Second, it combines models from multiple premises using a procedure akin to taking the Cartesian product of a set of possibilities. Third, the algorithm reduces the model and checks it against models of the causal relations specified by the model theory. If an adequate match is found, the system generates the corresponding causal relation as a conclusion.

No prior computational cognitive theory explains how people infer causal relations from sets of causal premises. But, the algorithm can be improved further. As Figure 1 shows, many discrepancies exist between the algorithm's predictions and human participants' tendency to make certain causal inferences. For instance, the algorithm predicts that humans should frequently infer that *X prevents Z* from the following premises: *X causes Y* and *Y causes ¬Z*. But people seldom ever make such a response. Perhaps they operate on a different sort of inferential heuristic, or perhaps they deliberate on their initial inferences and consider multiple models consistent with the premises (see, e.g., Khemlani & Johnson-Laird, 2016). The present algorithm can serve as a foundation for causal reasoning systems that take such deliberations into account.

References

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

Battaglia, P. W., Hamrick, J. B., & Tenenbaum, J. B. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*, 110.

Beebee, H. (2004). Causing and nothingness. In J. Collins, N. Hall, and L. A. Paul (Eds.), *Causation and Counterfactuals*, Cambridge, MA: The MIT Press.

Bello, P., & Khemlani, S. (2015). A model-based theory of omissive causation. In *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Bello, P., Wasylyshyn, C., Briggs, G., & Khemlani, S. (2017). Contrasts in reasoning about omissions. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Bernstein, S. (2014). Omissions as possibilities. *Philosophical Studies*, 167.

Byrne, R. (2005). *The rational imagination: How people create alternatives to reality*. Cambridge, MA: The MIT Press.

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

Cummins, D. D., Lubart, T., Alksnis, O., & Rist, R. (1991). Conditional reasoning and causation. *Memory & Cognition*, 19.

Dowe, P. (2001). A counterfactual theory of prevention and'causation'by omission. *Australasian Journal of Philosophy*, 79.

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

Goodwin, G.P., & Johnson-Laird, P.N. (2005). Reasoning about relations. *Psychological Review*, *112*.

Hall, N. (2004). Two concepts of causation. *Causation and counterfactuals*.

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

Johnson-Laird, P.N. (2006). *How we reason*. NY: Oxford University Press.

Johnson-Laird, P.N., & Byrne, R.M.J. (1991). *Deduction*. Hillsdale, NJ: Erlbaum.

Johnson-Laird, P.N., & Khemlani, S. (2017). Mental models and causation. *Oxford Handbook of Causal Reasoning*, 1-42.

Johnson-Laird, P. N., Khemlani, S., & Goodwin, G.P. (2015). Logic, probability, and human reasoning. *Trends in Cognitive Sciences*, 19.

Kahneman, D., & Tversky, A. (1982). The simulation heuristic. In D. Kahneman & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases (pp. 201-208). Cambridge, England: Cambridge University Press.

Khemlani, S., Barbey, A., & Johnson-Laird, P.N. (2014). Causal reasoning with mental models. *Frontiers in Human Neuroscience,* 8.

Khemlani, S., Byrne, R. M., & Johnson-Laird, P. N. (2018). Facts and Possibilities: A Model-Based Theory of Sentential Reasoning. *Cognitive Science*, *42*, 1887-1924.

Khemlani, S., & Johnson-Laird, P. N. (2011). The need to explain. *Quarterly Journal of Experimental Psychology*, 64.

Khemlani, S., & Johnson-Laird, P.N. (2013). The processes of inference. *Argument & Computation*, 4, 1-20.

Khemlani, S., & Johnson-Laird, P.N. (2016). How people differ in syllogistic reasoning. In *Proceedings of the 38th Annual Meeting of the Cognitive Science Society*. Austin, TX: Cognitive Science Society.

Khemlani, S., & Johnson-Laird, P. N. (2017). Illusions in reasoning. *Minds and Machines*, 27.

Khemlani, S., Lotstein, M., Trafton, J.G., & Johnson-Laird, P.N. (2015). Immediate inferences from quantified assertions. *Quarterly Journal of Experimental Psychology*, 68.

Khemlani, S., Mackiewicz, R., Bucciarelli, M., & Johnson-Laird, P.N. (2013). Kinematic mental simulations in abduction and deduction. *Proceedings of the National Academy of Sciences*, 110.

Khemlani, S., Orenes, I., & Johnson-Laird, P.N. (2012). Negation: a theory of its meaning, representation, and use. *Journal of Cognitive Psychology*, 24.

Khemlani, S., Wasylyshyn, C., Briggs, G., & Bello, P. (2018). Mental models of omissive causation. *Memory & Cognition*, 46.

Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*.

Pearl, J. (2009). *Causality*. Cambridge University Press.

Peirce, C.S. (1931-1958). *Collected papers of Charles Sanders Peirce*. 8 vols. C. Hartshorne, P. Weiss, and A. Burks, (Eds.). Cambridge, MA: Harvard University Press.

Rottman, B. M., & Keil, F. C. (2012). Causal structure learning over time: Observations and interventions. *Cognitive psychology*, 64.

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

Stephan, S., Willemsen, P., & Gerstenberg, T. (2017). Marbles in Inaction: Counterfactual Simulation and Causation by Omission. In *Proceedings of the 39th Annual Meeting of the Cognitive Science Society*. London, UK: Cognitive Science Society.

White, P. A. (2014). Singular clues to causality and their use in human causal judgment. *Cognitive Science*, 38.

Wolff P., Barbey A., Hausknecht M. (2010). For want of a nail: how absences cause events. *Journal of Experimental Psychology: General*, 139.

Wolff, P., & Barbey, A. K. (2015). Causal reasoning with forces. *Frontiers in Human Neuroscience*, *9*, 1.