Acta Psychologica 209 (2020) 103139

Contents lists available at ScienceDirect

Acta Psychologica

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

Explanatory completeness☆

Joanna Korman[a,*], Sangeet Khemlani[b]

[a] Bentley University, USA
[b] Naval Research Laboratory, USA

ARTICLE INFO

Keywords:
Explanatory reasoning
Incompleteness
Causal reasoning
Mental models

ABSTRACT

All explanations are incomplete, but reasoners think some explanations are more complete than others. To explain this behavior, we propose a novel theory of how people assess explanatory incompleteness. The account assumes that reasoners represent explanations as causal mental models – iconic representations of possible arrangements of causes and effects. A complete explanation refers to a single integrated model, whereas an incomplete explanation refers to multiple models. The theory predicts that if there exists an unspecified causal relation – a gap – anywhere within an explanation, reasoners must maintain multiple models to handle the gap. They should treat such explanations as less complete than those without a gap. Four experiments provided participants with causal descriptions, some of which yield one explanatory model, e.g., *A causes B* and *B causes C*, and some of which demand multiple models, e.g., *A causes X* and *B causes C*. Participants across the studies preferred one-model descriptions to multiple-model ones on tasks that implicitly and explicitly required them to assess explanatory completeness. The studies corroborate the theory. They are the first to reveal the mental processes that underlie the assessment of explanatory completeness. We conclude by reviewing the theory in light of extant accounts of causal reasoning.

1. Introduction

Suppose that a man begins to sneeze on a hike through the woods. Here is one explanation for his experience:

1a. Being outside caused him to breathe in pollen.
b. Breathing in pollen caused sneezing.

The explanation may seem acceptable because it provides a sensible sequence of events to explain the sneezing. But, you can elaborate it further: you may hypothesize why the man was outside in the first place. Human curiosity may depend on the realization that a putative explanation is incomplete.

Some philosophers of science hold that the notion of a "complete" explanation is nonsensical. Hempel, for instance, observed that scientists should judge an explanation to be complete "only if an explanatory account…had been provided for all of its aspects", but that the notion of completeness was "self-defeating" because any explanation can have "infinitely many aspects" (Hempel, 1965/2002). Other theorists concur: for instance, Rescher argued that "the finitude of human intellect" demands that we do not equate the adequacy of an explanation with how complete it is (Rescher, 1995, p. 8; see also Josephson, 2000; Railton, 1981, p. 239).

However, while explanatory completeness is an intractable notion in the abstract, the finitude of human intellect does not prevent reasoners in daily life from judging that some explanations are more complete than others. Pioneering work by Miyake (1986) showed that when people try to explain a particular phenomenon (e.g., how a sewing machine works), they often alternate between feeling, on the one hand, that their understanding of the phenomenon is satisfactory and complete, and on the other, that their understanding needs elaboration. In Miyake's studies, there often came a "bottoming out" point at which reasoners found it too difficult to elaborate on the mechanisms of sewing machines any further, either because the relevant information was uncertain or unavailable, or because they failed to recognize what they did not know. Psychologists such as Keil (2006) argue that the overwhelming complexity of the world puts highly detailed explanations of phenomena beyond the reach of individuals, and so people have no choice but to get by with incomplete, partial explanations that they nevertheless construe as complete.

More recently, Zemla et al. (2017) asked participants to evaluate natural explanations. They found that assessments of an explanation's incompleteness predicted judgments of that explanation's quality –

☆ Portions of this manuscript previously appeared in the 2018 *Proceedings of the Cognitive Science Society*. The work was funded by a National Research Council Research Associateship (to J.K.) and funding from the Office of Naval Research (to S.K.)

* Corresponding author at: Bentley University, 175 Forest Street, Waltham MA 02452, USA.
E-mail address: jkorman@bentley.edu (J. Korman).

https://doi.org/10.1016/j.actpsy.2020.103139
Received 30 August 2019; Received in revised form 16 June 2020; Accepted 12 July 2020
Available online 01 August 2020
0001-6918/

when people judged an explanation to be incomplete, they construed it as a bad explanation ($r = -0.65$). As Zemla et al.'s analysis suggests, explanatory completeness and quality can, at times, diverge: explanations can be complete but unconvincing. Scientific investigations routinely yield convincing but tentative, incomplete explanations, i.e., those that explain available facts but whose internal mechanisms leave relevant causal relations unspecified. Contemporary astronomers, for instance, posit the existence of an as-yet-unobserved ninth planet to explain why the Solar System wobbles away from its center (Batygin & Brown, 2016). Such an explanation is "good" insofar as it accounts for many different observations, but it's incomplete without an articulation of what the planet is made of and how it affects its nearest celestial bodies. Without specifying these and other relevant facts, the causal relation between the explanation (the planet's existence) and the relevant phenomenon in question (the wobbling of the Solar System) remains underspecified; that is, the explanation is incomplete. Assessments of completeness can therefore diverge from assessments of quality: an explanation's quality depends on corroboratory evidence, while an explanation's completeness depends on specifying the relevant causal relations between it and the phenomenon in need of explanation.

As Hempel, (1965/2002), Keil (2006), and others observe, people can elaborate on causal explanations ad infinitum because they can infer earlier and earlier root causes. For example, you could generate an explanation for why a person wanted to go on a hike (e.g., *perhaps he's looking for tranquility*), then you could generate an explanation for why he's looking for tranquility, and so on. Hence, people are justified in construing any explanation as incomplete. The present discussion focuses instead on how people make systematic judgments of completeness. These judgments may reflect reasoners' assessments of where gaps emerge in the specification of an explanation, and they may help explain the processes by which reasoners generate explanations (Horne et al., 2019).

To assess whether an explanation is complete or not, reasoners must mentally represent its components and assess its structure for unspecified causal relations, i.e., causal gaps. If the process discovers causal gaps, then the explanation should be deemed incomplete; otherwise, it should be deemed complete. Consider the explanations for the initial example in (2a) and (2b) below:

2a. An interest in tranquility caused a person to go outside.
b. Being outside caused the person to breathe in pollen, and that caused sneezing.

In isolation, (2a) seems like an incomplete explanation for why somebody sneezed outside. Incompleteness judgments may call on reasoners to consider background knowledge – i.e., reasoners may already know, based on a weather report, that being outside would cause pollen intake – but, in the absence of such specific knowledge, they should judge (2a) to be incomplete. In contrast, (2b) seems more complete – it is analogous to (1) above – because it provides the connection between being outside and sneezing. As far as we know, no studies have examined what kinds of explanations people consider complete.

In what follows, we present a novel theory that accounts for the mental representations that permit the rapid, online evaluation of explanatory completeness. The theory is based on the idea that humans build small-scale mental simulations – mental models – when they reason. Its central prediction is that reasoners should systematically judge an explanation to be incomplete when they are unable to build an integrated explanatory mental model of a phenomenon. We highlight several corollary predictions of the account, and then describe four novel experiments that tested and validated those predictions. We conclude by summarizing how extant theories of causal reasoning might be extended to deal with incomplete explanations.

1.1. A model-based theory of explanatory completeness

We present a novel account of explanatory completeness based on the assumption that people reason by constructing and inspecting small-scale mental simulations of possibilities. The theory abides by the constraints of the mental model theory of reasoning – the "model" theory, for short – which applies to reasoning in a variety of domains, including explanatory reasoning (Johnson-Laird et al., 2004; Khemlani & Johnson-Laird, 2011, 2012), and reasoning about causal, spatio-temporal, and abstract relations (Goldvarg & Johnson-Laird, 2001; Goodwin & Johnson-Laird, 2005; Kelly et al., under review). The theory makes three central claims:

- First, mental models are sets of iconic possibilities (Khemlani, Byrne, & Johnson-Laird, 2018). Iconicity implies that the structure of a model corresponds to the structure of what it represents (see Peirce, 1931-1958, Vol. 4). Hence, reasoners can represent causal sequences by mentally simulating the events in order in which they take place (Khemlani et al., 2014). Models are distinguished from mental imagery because they can represent abstract symbols, e.g., the symbol for negation (Khemlani et al., 2012).
- Second, reasoners distinguish *mental models* – which are initial, incomplete representations that represent only what is true of a given description – from *fully-explicit models* that represent both what is true and what is false in a given description. The theory posits two primary processes of inference: an intuitive construction process rapidly builds and scans initial mental models, but it is subject to various heuristics and biases. A slower, deliberative process revises the initial models into fully-explicit models, and it can eliminate systematic errors in reasoning (see, e.g., Khemlani & Johnson-Laird, 2017).
- Third, reasoners tend to be parsimonious: the more models that are required to solve a problem, the harder that problem will be, and most reasoners prefer to draw conclusions based on a single mental model.

The model theory explains how reasoners represent and make inferences from causal statements (Johnson-Laird & Khemlani, 2017; Khemlani et al., 2014), which underlie causal explanations. It posits that people understand causal statements as sets of possibilities. The meaning of a causal statement, such as *pollen causes sneezing*, refers to a set of three separate possibilities, depicted in this schematic diagram:

```
   pollen          sneezing   [cause and effect]
¬  pollen          sneezing   [no-cause and effect anyway]
¬  pollen       ¬  sneezing   [no-cause and no effect]
```

where '¬' denotes negation. Each row in the diagram represents a different temporally ordered possibility, e.g., the first row represents the possibility in which the cause and effect both occur, i.e., the possibility in which the person inhales pollen and then starts sneezing. The statement rules out the situation in which somebody inhales pollen and sneezing does not occur – and so the possibility is not included in the rows above. Hence, if it's true that *pollen causes sneezing*, there should not be any situation in which pollen is present but sneezing doesn't occur. If such a situation does occur, it acts as a counterexample to *pollen causes sneezing* – and a single counterexample can prompt reasoners to abandon, reject, or lower their belief in the relation (Frosch & Johnson-Laird, 2011). Some theorists counter that causal statements are probabilistic (e.g., Ali et al., 2011), i.e., reasoners should treat *pollen causes sneezing* as: *probably, pollen causes sneezing*. Recent evidence suggests that they should distinguish statements that include probabilistic qualifiers from those that don't (Goodwin, 2014), a result that contravenes the probabilistic account (see also Khemlani et al., 2014). And, as Pearl (2009) points out, conditional probabilities alone cannot distinguish statements such as *pollen causes sneezing* and *sneezing causes*

pollen. The model theory argues that reasoners represent causes by representing possibilities, not probabilities.

A strong prediction of the model theory is that when prompted to list the possibilities consistent with a given causal statement, reasoners should list the three possibilities above. Several laboratory studies corroborate the prediction (e.g., Goldvarg & Johnson-Laird, 2001; Khemlani, Wasylyshyn, et al., 2018). In daily life, however, reasoning based on the full meanings of causal statements demands significant cognitive resources to maintain each possibility in working memory. So, instead, reasoners rely on a single mental model to represent *pollen causes sneezing*:

pollen sneezing

A single model permits rapid inferences, since reasoners need to maintain only one possibility in memory. It also permits the rapid construction of a causal sequence of events. The theory posits, for instance, that when reasoners comprehend the causal statement in (2b), they should build an initial model of the first clause, e.g.,

outside breathe-in-pollen

and then they should combine the contents of the second clause with the model of the first, e.g.,

outside breathe-in-pollen sneezing

to create a single model of the phenomenon. One advantage to representing the causal sequence as a single possibility is that reasoners can scan the possibility to rapidly draw temporal inferences, e.g.,

3a. The person breathed in pollen before sneezing.
b. The person sneezed after he went outside.

The inference in (3a) comes about as a result of scanning the possibility from the earliest to the most recent event in the sequence. The inference in (3b) reflects a scan of the possibility in the opposite direction.

Explanations are often causal in nature (but not always; see Khemlani, 2018). Hempel and Oppenheim (1948) distinguished two elements of any explanation: the phenomenon to be explained (which they referred to as the *explanandum*) and the set of propositions that serve as explanations of the phenomenon (which they referred to as the *explanans*). The philosophers were interested in scientific explanations, but psychologists have applied it to everyday explanations elicited by naïve reasoners. The model theory posits that an explanatory mental model is one that represents an explanans and an explanandum, as in the model below:

outside breathe-in-pollen **sneezing**

In the model, sneezing serves as the explanandum and the two events that precede it serve as the explanans. Hence, an explanatory mental model is an efficient way to represent an explanation: it permits reasoners to draw a variety of relevant temporal and causal inferences.

What makes an explanation complete? To account for how people detect incompleteness, we posit the following hypothesis:

> **The completeness hypothesis.** A *complete* causal explanation is an explanatory mental model – a model that contains both an explanandum and an explanans – of a single possibility. Incomplete explanations are those that either do not represent an explanans, do not represent an explanandum, or represent two or more possibilities that may or may not have events in common. Reasoners should spontaneously construct and prefer complete explanatory models to incomplete ones.

Table 1
Mental models for complete and incomplete explanations.

Explanatory completeness	Verbal description	Mental model(s)	# of models
Complete	A causes C.	A C	1
Complete	A causes B. B causes C.	A B C	1
Incomplete	A causes B.	A B	2
	A causes C.	A C	
Incomplete	A causes C.	A C	2
	B causes C.	B C	
Incomplete	A causes X.	A X	2
	B causes C.	B C	

The hypothesis explains why (2a) above should be considered incomplete: its mental model does not represent an explanandum;

tranquility outside

The hypothesis also predicts that the explanation for sneezing in (4) below should be considered incomplete:

4. Hearing an alarm caused a person to be outside, and having a cold caused sneezing.

Reasoners must represent the two clauses with two separate models, e.g.:

alarm outside
cold sneezing

The model represents an explanans – the cold – as well as an explanandum – the sneezing. But it represents two possibilities instead of one, and so (4) alone cannot yield an integrated model. Some reasoners may spontaneously consult background knowledge to reconcile the two possibilities, but, absent such reconciliation, the two possibilities refer to an incomplete explanation.

Complete and incomplete explanations can produce several typical sorts of models, and Table 1 provides a list of them. The first two rows of the table show that the theory allows for the indefinite elaboration of complete explanations: for any complete model (of, e.g., *A causes C*) reasoners can integrate additional knowledge of the intervening causes that lead from a cause to an effect without impacting the model's completeness (e.g., *A causes B* and *B causes C*). Of course, it may well be the case that people consider elaborate explanations—those with more intervening causes—more complete than explanations with fewer intervening causes. While no other theory of explanatory completeness exists, we suspect that any potential account of how reasoners detect incompleteness should make the same prediction, and so it is not necessarily diagnostic of the model theory. Hence, in Table 2, we outline four predictions unique to the model theory.

First, the model theory predicts that in some cases, reasoners should be unable to combine separate sets of possibilities to yield an integrated explanatory model, such as when constructing a causal sequence of events. Preliminary evidence in support of the idea comes from Hegarty (2004), who showed that reasoners understand mechanical and physical devices in a piecemeal fashion, i.e., they consider the interdependencies between their internal components one after the other, and they have difficulty mentally integrating all but the simplest types of physical systems. In many cases, reasoners should have difficulty building and maintaining even simple networks of interconnected causal relations. Consider various explanations for a headache in

Table 2
Corollaries of the completeness hypothesis, specific predictions of the corollary, and a list of experiments below that tested the prediction.

#	Corollary	Prediction	Experiment
1	Reasoners should consider explanations in the form of a causal chain to be more complete than those in the form of "common cause" and "common-effect" relations.	Reasoners should judge (i) complete more often than (ii): i. A causes B. B causes C. ii. A causes C. B causes C.	1
2	Reasoners should consider causal chains as more complete than "diamond" sequences, even though both have only one root cause.	Reasoners should judge (i) complete more often than (ii): i. A causes B. B causes C. C causes D. D causes E. ii. A causes B. A causes C. B causes D. C causes D.	2
3	Reasoners should spontaneously distinguish completeness f incompleteness in explanations as a function of the number of models the explanation yields.	Reasoners should judge the explanation *A leads to C* as a more accurate summary of (i) than of (ii): i. A causes B. B causes C. ii. A causes X. B causes C.	3
4	Reasoners should seek out information that could potentially reconcile multiple explanatory models into a single, integrated explanatory model.	Reasoners should ask for information about *B* more often for (ii) than for (i): i. A causes B. B causes C. ii. A causes X. B causes C.	4

examples (5a-c) below:

5a. Breathing in pollen caused sneezing, and sneezing caused a headache. [causal chain]
b. Breathing in pollen caused sneezing, and breathing in pollen caused a headache. [common cause]
c. Breathing in pollen caused a headache, and having a cold caused a headache. [common effect]

The completeness hypothesis predicts that people should judge (5a) as more complete than (5b) or (5c), since only (5a) yields an integrated model of the three events in the premises. The model theory treats both "common cause" (5b) and "common effect" (5c) explanations as incomplete (cf. Read, 1988; Rehder & Hastie, 2004; Salmon, 1978). Other theories treat causal relations as graphical networks (e.g., Ali et al., 2011; Rips, 2010; Sloman et al., 2009) where a set of nodes represents causal events, and directed connections between nodes represent causal relations (see Fig. 1). No causal network account makes any prediction about completeness judgments, but in principle, such theories may posit that reasoners should make completeness judgments based on the completeness of a network. A network may be considered complete if it can be represented as a connected graph, i.e., a graph in which there exists at least one path between any two nodes. Such an account would make no such distinction between (5a) and (5b and c), because all three cases can be modeled by connected graphical networks.

For the same reason, theories based on causal networks should assume that reasoners treat the two causal structures depicted in Fig. 1 as analogously complete. The figure depicts a graph representation of a chain structure (panel 1) as well as a "diamond" structure, both of which contain four interconnected nodes. Since both types of structures yield connected graphs, theories based on causal networks should treat as complete any explanations that mimic their structures. Other accounts of explanatory reasoning base explanatory preferences on "root causes" (e.g., Pacer & Lombrozo, 2017); such accounts would likewise treat both structures in Fig. 1 as complete, because they both have a single root cause (node A). The model theory, in contrast, treats causal sequences (panel 1) as complete but diamond sequences (panel 2) as incomplete (see Table 2).

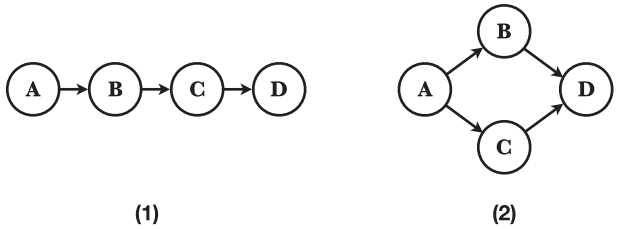

Fig. 1. Causal graphical network depictions of a causal chain structure (panel 1) and a "diamond" structure (panel 2). Both networks are integrated, i.e., such that each node in the network is connected to at least one other node. And both networks have one root cause (node A) and one resulting effect (node D).

Third, the completeness hypothesis predicts that reasoners should prefer complete explanations to incomplete explanations on tasks that do not explicitly require them to think about completeness (cf. Zemla et al., 2017). One such task is to assess whether a particular causal explanation is an accurate summary of a causal description of events. Consider (6a) and (6b) below:

6a. Being outside caused a person to breathe in pollen.
Breathing in pollen caused sneezing.
b. Being outside caused a person to breathe in pollen.
Neural activity caused sneezing.

The theory predicts that the following statement:

Being outside leads to sneezing.

is an accurate summary of (6a) but not (6b). The former set of causal statements is compatible with a single explanatory model, but the latter is not, because (6b) does not make explicit how breathing in pollen relates to neural activity.

Finally, the theory predicts that when reasoners seek out additional information, they should do so relative to the information missing in their explanatory mental models. For instance, when given the opportunity, reasoners should wonder what *breathing in pollen* causes more often for (6b) than for (6a).

Four preregistered experiments tested the theory's predictions. In Experiments 1 and 2, participants directly compared the relative completeness of causal chain structures, common-cause and common-effect structures (Experiment 1), and diamond structures (Experiment 2). Experiments 3 and 4 assessed the way reasoners detect explanatory incompleteness in more implicit ways: Experiment 3 asked participants to judge the accuracy of summary conclusions for complete and incomplete descriptions, and Experiment 4 asked participants to select one or more events in an incomplete causal chain to investigate further. All four studies corroborated the predictions outlined in Table 2.

2. Experiment 1

Experiment 1 explored the conditions under which reasoners directly evaluate an explanation as complete or incomplete, i.e., it tested the theory's first prediction: reasoners should consider explanations in

the form of causal chains (e.g., *A causes B* and *B causes C*) more complete than those including the same three components, but taking the form of common-cause structures (e.g., *A causes B* and *A causes C*) or common-effect structures (e.g., *A causes C* and *B causes C*). The experiment served as a strong test of the completeness hypothesis: if an explanation's perceived completeness depends on the number of possibilities represented, then they should judge causal chains to be more complete than common-cause or common-effect structures.

2.1. Method

2.1.1. Participants

50 participants completed the experiment for monetary compensation through Amazon Mechanical Turk. All of the participants were native English speakers, and all but 8 had taken one or fewer courses in introductory logic.

2.1.2. Preregistration and data availability

The predicted effects were pre-registered through the Open Science Framework platform. The data, analysis scripts, preregistrations, and experimental code for Experiment 1 and all subsequent experiments are available at: https://osf.io/s6fb7/.

2.1.3. Design and procedure

The experiment invited participants to think of themselves as teachers who had to evaluate their students' notes on novel phenomena. The notes consisted of causal explanations for an event, *C*, that linked several events together. Participants carried out 8 problems altogether. On half of the problems, participants received explanatory causal chains in the following schematic:

A causes B.
B causes C.

where *A, B,* and *C* stood for various properties and behaviors of imaginary entities. The theory predicts that reasoners should build a single integrated mental model of the explanation above, e.g.,

A B C

The remaining half of the problems presented participants with problems that the theory construed as incomplete. 2 of those problems made use of explanations for an effect, *C*, with a common-cause structure:

A causes B.
A causes C.

where the causal relation between *B* and *C* is unspecified. And the other 2 made use of a common-effect structure, e.g.,

A causes C.
B causes C.

where the relation between *A* and *B* went unspecified. The theory predicts that reasoners should be unable to construct an integrated mental model from common-cause or common-effect explanatory structures, and so they should be judged relatively less complete.

For each problem, participants evaluated whether an explanation for event *C* (the explanandum) was complete by using a slider bar to indicate a number on a Likert scale from 1 (definitely incomplete) to 5 (definitely complete). The slider bar's default position was set to 3 (I cannot be certain), but participants were not permitted to proceed to the next trial until they shifted the slider bar from the default position to some other position. This constraint served as an attention-check mechanism: it sought to prevent participants from answering without processing the meaning of the text. The instructions presented participants with an example of set of notes for a fictitious phenomenon – i.e., notes that explained why a species of lizard called a *Sclerdid* is a good swimmer – and then provided them with the following instruction:

> *Your task will be to assess whether the student's notes explaining why* Sclerdids are good swimmers *are complete, or whether the notes are missing one or more pieces of information (and are incomplete).*

Participants acted as their own controls and carried out all 8 problems in a fully repeated measures design. They completed two practice trials (one yielding a single mental model, another requiring multiple models), and they received the rest of the problems in a randomized order.

2.1.4. Materials

Materials were drawn from four separate domains (natural, biological, social, and mechanical). Each set of materials was a collection of candidate causal events that concerned properties or behaviors of an imaginary entity. These properties and behaviors were designed such that any one property or behavior could serve as a cause or a resulting effect of any other (see Appendices A and B). Participants received premises in the following format:

[Event A] causes the Zindo to [event B].
[Event B] causes the Zindo to [event C].

where events *A*, *B*, and *C* were assigned randomly from a pool of five candidate events. For instance, one set of materials concerned a mechanical device used in factories called a “Zindo,” and so some participants may have received the following set of premises:

Releasing a valve causes the Zindo to engage a pump.
Narrowing an aperture causes the Zindo to engage a pump.

Participants assessed whether the premises provide a complete or an incomplete explanation for, e.g., why the Zindo engages a pump (the explanandum). For each problem, the experiment randomly assigned the three events (releasing valve, narrowing an aperture, and engaging a pump) to events *A*, *B*, and *C* according to the three types of problems in the study. The contents were rotated around the conditions so that only one problem in the study was assigned to a particular condition – for instance, the materials describing the “Zindo” appeared as only a complete description or else as an incomplete description for a particular participant. Across the study as a whole, that material appeared the same number of times in complete or incomplete conditions. For examples of problems presented to participants in Experiment 1 and all subsequent experiments, see Appendix B.

2.2. Results and discussion

Fig. 2 presents the mean completeness ratings participants gave for each of the three types of problems presented in Experiment 1. Participants rated causal chains as more complete ($M = 3.26$) than common-cause structures ($M = 2.64$, Wilcoxon test, $z = 3.19$, $p = .001$, Cliff's $\delta = 0.62$) or common effect structures ($M = 2.83$, Wilcoxon test, $z = 2.05$, $p = .04$, Cliff's $\delta = 0.43$). The predicted pattern of responses was robust, and additional experiments (not reported here) replicated the effect. A parallel analysis excluded participants with outlying reaction times – it revealed similar effects for Experiment 1 and all further experiments, and so we omit those analyses for brevity.

One concern with the nonparametric analyses is that they do not control for the variance contributed by the materials or the individual participants. So, we ran a generalized linear mixed model regression that treated participants' judgments of completeness as the outcome variable and the three types of problem as a fixed effect; it controlled for material and participant noise. The analysis revealed that the

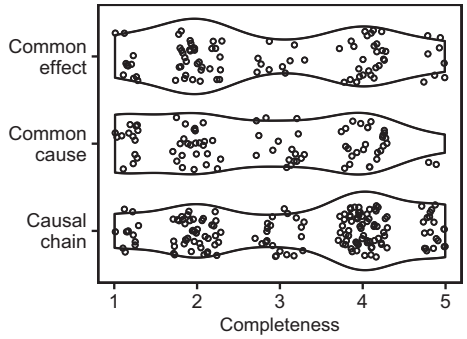

Fig. 2. Density plots of participants' responses to the three conditions in Experiment 1; the width of each shape is proportional to participants' response frequencies.

problem types reliably predicted judgments of completeness ($\beta = -0.49$, $p < .0001$), further corroborating the theory's first prediction.

The distribution of participants' responses exhibited a bimodal pattern (Hartigan's dip test, $D = 0.13$, $p < .0001$): participants tended not to endorse the midpoint of the scale (likely because of the attention-check mechanism the experiment implemented), and instead preferred to select the upper and lower values of the scale. For complete problems, they selected the upper values of the scale (4 and 5) 54% of the time. For incomplete problems, they selected the upper values of the scale 36% of the time (Wilcoxon test, $z = 3.59$, $p = .0003$, Cliff's $\delta = 0.18$).

In sum, Experiment 1 corroborated the first prediction of the completeness hypothesis: participants penalized common-cause and common-effect structures, i.e., they evaluated them as relatively less complete. Experiment 2 sought to test the theory's second prediction.

3. Experiment 2

Experiment 2 was similar to Experiment 1, but it included materials designed to test the theory's second prediction: reasoners should consider explanations in the form of causal chains to be more complete than those explanations in the form of diamond structures (see Fig. 1 and Table 2). Participants in the study therefore evaluated two types of explanation: the first described a causal chain, and the second described a diamond structure. A preference for causal chains over diamond structures suggests that reasoners base completeness judgments on the number of possibilities represented, and it rules out the hypotheses that reasoners base completeness judgments on network structures or on the number of root causes in an explanation.

3.1. Method

3.1.1. Participants

51 participants completed the experiment for monetary compensation through Amazon Mechanical Turk. All of the participants were native English speakers, and all but 5 had taken one or fewer courses in introductory logic.

3.1.2. Preregistration and data availability

The predicted effects were pre-registered through the Open Science Framework platform.

3.1.3. Design and procedure

The experiment used a design, procedure, and task identical to Experiment 1. Participants carried out 8 problems altogether. On half the problems, participants received explanatory causal chains in the following schematic:

A causes B.
B causes C.
C causes D.
D causes E.

where *A, B, C, D,* and *E* stood for various properties and behaviors of imaginary entities. The theory predicts that reasoners should build a single integrated mental model of the explanation above, e.g.,

A B C D E

The other half of the problems presented participants with a diamond structure, which the theory treats as an incomplete explanation. Premises conformed to the following schematic:

A causes B.
A causes C.
B causes D.
C causes D.

The completeness hypothesis predicts that people should represent two possibilities when constructing a model of the premises above:

A B D
A C D

A network-based theory, however, would predict that people should construct a single network integrating all four premises (see Fig. 1, panel 2). As in the previous study, participants were provided an explanandum, i.e., event *D*, and evaluated whether the explanation was complete by using a slider bar to indicate a number on a Likert scale from 1 (definitely incomplete) to 5 (definitely complete). As in the previous study, the default position of the slider on each trial was set to 3 (I cannot be certain), and participants were required to modify the default position to select a response. The experiment implemented a fully repeated measures design using the same programming framework as in Experiment 1. Participants completed two practice trials (one yielding a single mental model, another requiring multiple models), and they received the rest of the problems in a randomized order.

3.1.4. Materials

The materials were the same as those used in Experiment 1. However, Experiment 1 made use of only 3 of the 5 available causes for each material (see Appendix A). Experiment 2 made use of all 5 of the available causes.

3.2. Results and discussion

Fig. 3 presents the mean completeness ratings participants gave for the 2 types of problems in Experiment 2. Participants provided higher ratings for causal chains, which can be represented with a single explanatory mental model ($M = 3.35$), than for diamond structures ($M = 2.86$, Wilcoxon test, $z = 3.88$, $p = .0001$, Cliff's $\delta = 0.20$). A generalized linear mixed model regression analysis that treated participants' judgments of completeness as the outcome variable, the types of problem as a fixed effect, and material and participants as random effects revealed that the problem types reliably predicted judgments of completeness ($\beta = -0.41$, $p = .0004$), further corroborating the completeness hypothesis's second prediction.

As in the previous study, the attention-check mechanism may have reduced participants' tendency to select midpoint responses, and so their responses exhibited a bimodal pattern (Hartigan's dip test, $D = 0.13$, $p < .0001$). For causal chains problems, they selected the upper values of the scale (4 and 5) 50% of the time. For diamond

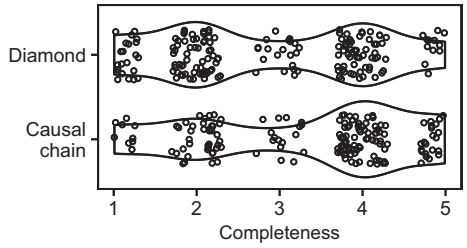

Fig. 3. Violin plot of participants' responses to diamond structure problems and causal chain problems in Experiment 2. The width of each shape is proportional to participants' response frequencies.

structures, they selected the upper values of the scale 41% of the time (Wilcoxon test, $z = 3.64$, $p = .0003$, Cliff's $\delta = 0.18$).

Experiments 1 and 2 corroborated the first and second predictions of the completeness hypothesis (see Table 2): reasoners judged explanations more complete when their premises could be integrated into a single mental model compared to sets of premises that yield multiple models. Nevertheless, participants may have distinguished between complete and incomplete explanations in accordance with the model theory only because the task prompted them to consider an explanation's completeness. Accordingly, rather than soliciting explicit judgments of completeness, Experiment 3 assessed whether reasoners spontaneously distinguish complete from incomplete explanations based on judgments of accuracy. Unlike Experiments 1 and 2, it also employed a design that varied the number of causal statements that participants considered.

4. Experiment 3

Experiment 3 tested the third prediction of the completeness hypothesis (see Table 2). The experiment presented participants with causal descriptions of events; it manipulated whether those descriptions concerned relations that could be integrated into a single model or whether the relations were systematically underspecified such that they could only be represented with multiple models. Unlike the previous studies, participants in Experiment 3 did not directly evaluate whether explanations were complete or incomplete. Instead, they received a description of causal relations and assessed the accuracy of a single-model explanation that summarized the description. The experiment also presented participants with descriptions of two different lengths.

4.1. Method

4.1.1. Participants

30 participants completed the experiment for monetary compensation through Amazon Mechanical Turk. All of the participants were native English speakers, and all but 2 had taken one or fewer courses in introductory logic.

4.1.2. Design and procedure

As in the previous study, Experiment 3 invited participants to think of themselves as teachers who had to evaluate their students' notes on novel phenomena. The notes consisted of a set of causal statements that linked novel events together. On half of the problems, participants received causal statements in the following schematic:

A causes B.
B causes C.
C causes D.

where *A, B, C,* and *D* stood for various properties and behaviors of imaginary entities. The participant then received a summary explanation, e.g.,

A leads to D.

The summary explanation served to link an initial event, i.e., event *A*, to the expanandum, i.e., event *D*. Participants were told that the student had drawn such an explanation from their notes. For each problem, participants evaluated whether the summary explanation was correct by clicking "Yes" or "No." The model theory predicts that most reasoners should build a single integrated mental model of the statements above, e.g.,

A B C D

Hence, they should be likely to judge the summary explanation as accurate.

The other half of the problems provided participants with a set of statements that could not be integrated into a single model, e.g.:

A causes X.
B causes C.
C causes D.

X is not a cause of *B*, *C*, or *D*, and so the theory predicts that reasoners will maintain separate models in order to represent the statements, e.g.:

A X
B C D

Half of the problems concerned two premises (*A caused B* and *B caused C*) and the other half concerned three (the previous two, plus *C caused D*). Hence, the length of the descriptions for one-model and multiple-model problems was held constant (see Appendix B).

Participants acted as their own controls and carried out 8 problems in a fully repeated measures design that manipulated whether the descriptions yielded one model or multiple models, and whether the problems comprised two premises or three. Participants completed 2 practice trials (a one-model problem and a multiple-model problem), and they received the remaining problems in a randomized order. The experiment randomized the positions of the "yes" and "no" buttons on the screen.

4.1.3. Materials

The experiment drew on the same materials used in Experiments 1 and 2. Each set of materials consisted of 5 separate events that could serve as both a cause and an effect. As before, for each problem, the experiment randomly assigned the materials to the different events in the explanation (e.g., *A*, *B*, *C*, *D* or *X*) such that every participant received a unique set of problem contents. The following is an example multiple-model problem:

Releasing a valve causes the Zindo to open a glass pane.
Engaging a pump causes the Zindo to flip a switch.
Flipping a switch causes the Zindo to narrow an aperture.

Those who received the problem had to evaluate the accuracy of the following summary explanation for why the Zindo narrows an aperture:

Releasing a valve leads the Zindo to narrow an aperture.

The summary explanation employed the construction: *A leads [the entity] to do D*, where event *A* is the initial event and event *D* is the explanandum. By evaluating the accuracy of this explanation as a summary of the causal description provided in the premises, reasoners indirectly indicated whether they detected an explanatory gap somewhere between *A* and *D*.

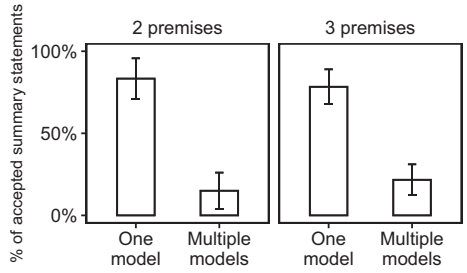

Fig. 4. Percentages of responses for which participants accepted the summary statements in Experiment 3 as a function of whether the description of events in the problem yielded one model or multiple models, and as a function of the number of premises in the description. Error bars show 95% confidence intervals.

4.2. Results and discussion

Fig. 4 presents the proportions of trials on which participants accepted the summary explanations in Experiment 3. Participants evaluated summary explanations as more accurate for one-model descriptions than for multiple-model descriptions (81% vs. 18%; Wilcoxon test, $z = 7.86, p < .0001$, Cliff's $\delta = 0.63$). The number of premises did not affect participants' evaluations, either overall (49% vs. 50%; Wilcoxon test, $z = 0.13$, $p = .90$, Cliff's $\delta = 0.01$), or for multiple-model problems in particular (15% vs. 22%; Wilcoxon test, $z = 1.07$, $p = .29$, Cliff's $\delta = 0.07$). A generalized linear mixed model regression controlled for material and participant noise; the analysis yielded a similar outcome, i.e., that the presence of a gap was the only significant main effect ($\beta = -4.89$, $p = .005$).

The results from Experiment 3 corroborated the third prediction of the completeness hypothesis: participants detected whether summary explanations were incomplete on the basis of whether they could be represented by a single model, or whether they required multiple-models. Two potential deflationary accounts might explain the finding. First, descriptions that yielded multiple models may have contained more words and syllables than those that yielded one model. However, the structure of the premises and the random assignment of materials to events prevented any such systematic confound. Indeed, the random assignment algorithm routinely generated one-model problems that contained more words than multiple-model problems. A more serious limitation in the design is that all multiple-model descriptions concerned one additional event – *X* – relative to one-model descriptions. However, if participants responded on the basis of the number of events in a problem, then their judgments of accuracy should differ between two-premise problems and three-premise problems. In fact, their judgments did not differ as a function of the number of events (see Fig. 3).

Other theories of causal reasoning may account for the results reported in Experiment 3 – for instance, a theory based on graphical networks may explain the results as a failure to build an integrated network instead of appealing to multiple models. Hence, the model theory's third prediction (see Table 2) is not uniquely diagnostic of the theory. But, no alternative account derives the third prediction based on the construction of mental models. We return to this issue by considering two alternative accounts – network-based theories and theories based on root-causes – in the General Discussion.

As one reviewer noted, the effects described in Experiment 3 may be trivial – that is, any theory of causal reasoning should explain them as a failure to make a transitive inference. The claim depends on the view that causation is a transitive relation, e.g., if *A causes B* and *B causes C*, then *A causes C*. Theorists have pointed out limitations in treating causation as transitive (Hall, 2000; McDermott, 1995). In certain contexts, reasoners appear to systematically reject causal transitivity (Johnson & Ahn, 2015). Moreover, the problems given to participants cannot be treated as logically valid or logically invalid, because some systems of causal logic permit transitive inferences and others do not. Nevertheless, people often accept transitive conclusions from causal premises (von Sydow, Hagmayer, & Meder, 2016), and the design of Experiment 3 cannot rule out the possibility that considerations of transitivity, not completeness, drove participants' behavior.

The value of explanations in daily life is that they are instructive: they serve as the basis for decisions, and they imply consequences. Suppose, for instance, that you turn a light switch on and observe that the bulb to which it's attached doesn't light up. At least three explanations seem reasonable: the bulb is out, the power is out, or the switch is malfunctioning. Your subsequent decisions – whether to change the lightbulb or call the power company – depend on your explanation. You might gather additional information to validate your initial explanation, e.g., you may unscrew the lightbulb under the assumption that the bulb is fused. The experiments reported thus far used tasks designed to capture reasoners' evaluations of explanations, but reasoners in daily life may use a gap in an explanation as the basis for investigating additional information. The completeness hypothesis predicts that reasoners should search for information that reconciles multiple explanatory models into one integrated model. Experiment 4 sought to test this final prediction of the hypothesis.

5. Experiment 4

Experiment 4 tested the fourth prediction of the completeness hypothesis, which states that when given the opportunity, reasoners should seek out additional information that allows them to build a single integrated explanatory mental model. In other words, their search for information should be deliberative – they should consider irrelevant any new information that does not help reconcile a set of models.

5.1. Method

5.1.1. Participants

30 participants from the same pool as the previous studies completed the experiment for monetary compensation. All of the participants were native English speakers, and all but three had taken one or fewer courses in introductory logic.

5.1.2. Materials and design

The materials and design were the same as in Experiment 3. Participants acted as their own controls and carried out 8 problems in a fully repeated measures design, which manipulated whether the descriptions referred to one model or multiple models. As in previous studies, materials were randomly assigned to the different events in the explanation such that every participant received a unique set of problems. Additionally, the study employed the same nested design used in Experiment 3: two of the multiple-model problems placed the gap in the first premise the other two placed the gap in the second premise. Participants completed two practice trials (one that yielded one model and one that yielded multiple models), and they received the problems in a randomized order.

5.1.3. Procedure

The procedure was similar to that used in Experiment 3. Experiment 4 presented participants with students' notes about imaginary entities and the students' conclusions from these notes; the conclusions took the same form as they had in previous studies (e.g., *A leads to D*). Unlike in Experiment 3, however, the participants' task was to select which one of the four candidate causal events (*A, B, C,* or *D*) to research further. The task is illustrated in the following example of a multiple-model problem:

Here are some research notes your student took about the Zindo, a mechanical device used in factory assembly lines:

- *Flipping a switch causes the Zindo to open a glass pane.* [*A causes X*]
- *Releasing a valve causes the Zindo to narrow an aperture.* [*B causes C*]
- *Narrowing an aperture causes the Zindo to engage a pump.* [*C causes D*]

From the above notes, your student concludes the following:

Flipping a switch leads the Zindo to engage a pump.

Given the conclusion the student has reached, which item from among those below should you tell your student to research further?

1. *What leads the Zindo to flip a switch.* [*A*]
2. *What leads the Zindo to release a valve.* [*B*]
3. *What leads the Zindo to narrow an aperture.* [*C*]
4. *What leads the Zindo to engage a pump.* [*D*]

Participants had to select an appropriate answer among options corresponding to the four events common to problems with and without gaps. The experiment randomized the order of the options.

5.2. Results

Figure 5 provides the proportions of participants' selections of causal events as a function of the different conditions in the study. Participants preferred to ask questions about distinct events in the causal description as a function of the presence of a gap, and as a function of the location of the gap within the description. The experiment measured their selections among the *A*, *B*, *C*, or *D* events, which were dummy-coded and subjected to multiple Friedman non-parametric analyses of variances. These tests served as omnibus assessments of differences among the three problem types: one model, multiple model (gap in the first premise, at the *B* event), and multiple model (gap in the second premise, at the *C* event). As predicted, there were differences among the three problem types in the endorsement of each of the respective gap events (Friedman tests, $\chi^2 > 7.35$, $ps \leq 0.03$). Participants also differed across the three problem types in their endorsement of *A*, the first event in the chain (Friedman test, $\chi^2 = 19.05$, $p < .0001$).

Beyond these omnibus tests, planned comparisons examined specific differences across conditions. Participants selected the first event in the chain overwhelmingly more often for one-model than multiple-model problems (70% vs. 36%; Wilcoxon test, $z = 5.74$, $p < .0001$, Cliff's $\delta = 0.34$). The pattern is sensible, because the task required participants to select 1 of the 4 events, and so they appeared to ask about the only event in the chain that lacked a causal antecedent, i.e., event *A* (Fig. 5).

In contrast, participants selected the middle options (*B* or *C*) more often for multiple-model problems than for one-model problems (43% vs. 14%; Wilcoxon test, $z = 5.01$, $p < .001$, Cliff's $\delta = 0.28$). For multiple-model problems that positioned the gap in the first premise (i.e., for the *B* event), participants selected the *B* event more often than the *C* event (30% vs. 8%; Wilcoxon test, $z = 2.71$, $p = .007$, Cliff's $\delta = 0.22$), and the opposite pattern held for multiple-model problems that positioned the gap in the second premise (i.e., at the *C* event; 37% vs. 10%; Wilcoxon test, $z = 3.02$, $p = .002$, Cliff's $\delta = 0.27$. The results corroborated the fourth prediction of the completeness hypothesis: reasoners' based their selections on their detection of an unspecified causal relation in the provided description. When a causal description lent itself to the construction of multiple explanatory models, participants sought additional information that them to reconcile the scenario into a single explanatory mental model.

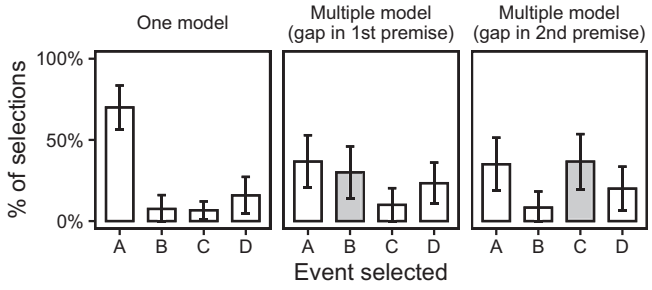

Fig. 5. Proportion of selections of causal events in the causal description as a function of whether the premises described a one-model problem, a multiple-model problem with a gap at the *B* event (highlighted in gray), or else a multiple-model problem with the gap at the *C* event (highlighted in gray).

6. General discussion

Reasoners think "complete" explanations are better than incomplete ones (Zemla et al., 2017). Early studies showed that construing an explanation as incomplete allows people to ask meaningful questions in order to fill in gaps in their understanding (Miyake, 1986). But on any objective notion of completeness, all explanations are incomplete (Hempel, 1965/2002). So what makes people judge an explanation as more or less complete? We developed an account that assumes that a complete explanation refers to causal representation of a single possibility, whereas an incomplete representation refers to multiple possibilities. We described a theory based on the view that reasoners construct mental models of causal relations to distinguish complete from incomplete explanations.

The theory makes four unique predictions: first, reasoners should consider explanations in the form of simple causal chains, e.g.,

A causes B and B causes C.

as more complete than explanations in the form of a common-cause structure, e.g.,

A causes B and A causes C

or a common-effect structure, e.g.,

A causes C and B causes C.

We conducted a study that tested and corroborated the prediction (Experiment 1). Second, reasoners should consider causal chains more complete than "diamond" structures, i.e., causal sequences that combine common-cause and common-effect structures; Experiment 2 corroborated the prediction. Third, reasoners should distinguish complete from incomplete explanations as a function of the number of models that a causal description yields. For instance, these two descriptions each yield one model:

A causes B.
A causes B and B causes C.

but this description yields multiple models:

A causes X and B causes C.

because reasoners cannot reconcile how *X* is related to *B* or *C*. Hence, reasoners should consider one-model problems complete but multiple-model problems incomplete. Experiment 3 tested and validated the theory's third prediction. The fourth and final prediction of the theory is that reasoners should seek out additional information in order to reconcile multiple models into a single explanatory model. Experiment 4 validated the prediction.

In what follows, we discuss criticisms of the completeness hypothesis as well as potential alternative accounts for the phenomena we describe. We conclude by describing how the present research can help in understanding how people generate explanations.

6.1. Criticisms of the completeness hypothesis

Experiments 1–4 finesse a dilemma in the evaluation of explanatory completeness: the infinite regress of root causes. In principle, as philosophers observe, all explanations are incomplete – but our studies sought to capture the more limited notion of people's subjective evaluations of completeness, and thus they made use of explanations with artificially fixed boundaries; to explain event *C*, for example, studies provided participants with causal relations in the form *A causes B* and *B causes C*, and so the explanation's root cause was fixed at, e.g., event *A*. If the experiments hadn't fixed the root cause, then some participants may have evaluated every explanation as incomplete – a justifiable response. Nevertheless, an open question concerns how and when reasoners choose to accept a proposed explanation or else infer a plausible mechanism that explains the events leading up to those described in the proposed explanation. Future studies should examine how reasoners generate explanations, and how they choose to stop the generative process (see, e.g., Horne et al., 2019).

One criticism of the results of these studies is that their results are all obvious: each experiment describes results that accord with intuitions. Hence, any theory of explanatory completeness – the model theory proposed here, or some other theory – should explain them. Readers who hold such a view may find it surprising that no extant theory of causal reasoning can explain the results from our studies, i.e., no alternative theory explains how or why reasoners should be sensitive to explanatory completeness. Yet, as Zemla et al. (2017) show, reasoners systematically rate some explanations as more complete than others, and those judgments predict participants' ratings of how good an explanation is. Without a theory of explanatory completeness, those patterns are inexplicable.

Another criticism may concern the first prediction of the completeness hypothesis, which states that common-cause arrangements are incomplete explanations. Common-cause explanations require reasoners to consider two different causes as separate possibilities, i.e., reasoners should interpret.

A causes B and A causes C.

as the following set of mental models:

```
A       B
A               C
```

because the premise leaves unspecified the relation between *B* and *C*. Skeptics may wonder if the treatment is sensible for all situations. As Salmon (1978) observed, it is reasonable for people to infer that a set of related events, such as the lights turning off at the same time in several homes, can be traced back to a common cause, such as a power outage (see also Reichenbach, 1956). The power outage explanation is sensible, but it would seem that the completeness hypothesis would treat it as yielding an incomplete set of models, i.e., a set of multiple models depicted in this diagram:

```
power-outage      blackout-home-1
power-outage      blackout-home-2
power-outage      blackout-home-3
                  ...
```

Each row represents the separate possibility in which the power outage leads to a blackout in a specific home. To cope with such situations, the theory could be extended to deal with common-cause explanations by allowing for models of generalized events, e.g.:

```
power-outage      blackout-home-x
```

Hence, an inductive strategy would combine the three models above to a model of a generalization, in which people represent a property that applies to a set of entities as a property that applies to a single representative entity that stands in place of the set:

```
blackout-home-1
blackout-home-2        →       blackout-home-x
blackout-home-3
```

Such a component yields a testable prediction: common-cause explanations should be more sensible for similar events (those that can be generalized) than dissimilar events. It also predicts that reasoners should take additional time to construct common-cause explanations – such explanations should be difficult to construct because of the need to deliberatively generalize across multiple events. The model theory is well-suited to explain both inferences about quantifiers (Khemlani, Lotstein, et al., 2015) as well as the distinction between the cognitive processes that underlie intuitive and deliberative processing (Khemlani, Byrne, & Johnson-Laird, 2018; Khemlani & Johnson-Laird, 2013). The extended account maintains the prediction of the completeness hypothesis that reasoners should attempt to coerce a set of multiple models into a single integrated model.

A special class of causation may challenge the present account: conjunctive causes. Conjunctive causes are situations in which an effect comes about from two independent events, neither of which is sufficient to bring about the effect on its own. For instance, consider the scenario in which two people work together to lift a heavy bookcase. One reviewer worried that the model theory would represent such causes as two separate models:

```
person-1                bookcase-lifted
          person-2      bookcase-lifted
```

As the reviewer observed, such an account would treat conjunctive causes as incomplete explanations. But the approach is inconsistent with the model theory at the outset, because the models above suggest that each individual can lift the bookcase independently, while they were intended to represent the scenario in which neither individual was sufficient. A better way to represent joint causation is to simulate both causes as a single scenario in a model such as the following:

```
[ person-1  person-2 ]      bookcase-lifted
```

where the brackets denote the joint interaction between two separate individuals to create a single event. This approach yields an integrated model of the situation, i.e., a complete explanation. An alternative approach may represent the sequence of events that unfold in time in a series of steps that mirror what would happen in the real world: the two individuals do not hold the bookcase; one person lifts one side; then the second person lifts the other side; then they both lift the bookcase together. Reasoners use kinematic simulations to solve problems and engage in abductive and deductive reasoning (Khemlani et al., 2013), and they can infer causal relations from such sequences (Khemlani, Goodwin, & Johnson-Laird, 2015). Kinematic simulations may explain how people interpret conjunctive causes.

6.2. Potential alternative theories and mitigating factors

Psychological theories of causal reasoning are not fixed in stone, and any of them can be amended to accept the four predictions we outline. Yet no account prior to the present one focused on how and why reasoners distinguish complete from incomplete explanations. Unlike other theories of causal reasoning, the model theory characterizes the increased representational burden that incomplete explanations impose on reasoners: it is more difficult for reasoners to represent multiple possibilities in mind and to draw inferences from multiple possibilities (Johnson-Laird, 1983; Khemlani & Johnson-Laird, 2017). The model theory thus makes the clear prediction that reasoners should detect incomplete explanations when they represent multiple possibilities.

Other accounts focus less on how reasoners maintain multiple representations; they depend instead on alternative mechanisms to explain causal inference. For instance, some theorists argue that causal reasoning requires the mental representation of causal Bayesian graphical networks (Ali et al., 2011; Rips, 2010; Sloman, 2005; Sloman et al., 2009). However, to explain the present data, any account that appeals to network representations would need to go beyond what current network-based theories propose. It would need to compare networks of varying sizes, and it would need to represent missing causal connections – gaps – between the nodes in a graph. Missing connections are difficult to express using causal networks, because networks typically treat gaps implicitly, i.e., as the absence of a connection between two nodes. And counting the number of absent connections for any non-trivial network leads to combinatoric explosion. The results we show in Experiment 4 demonstrate that people privilege some absent connections over others. Participants in that study made productive use of their detection of incompleteness: they were able to identify an event that could allow reasoners to integrate their incomplete representation. A network-based theory thus needs to explain how people identify privileged connections. Unlike causal networks, the present model-based theory allows for the direct representation of gap in a causal sequence – gaps yield multiple models – and makes predictions about the consequences of the increased representational burden.

It is possible to extend network-based theories with mechanisms that maintain, align, and compare multiple networks at a time. On such accounts, a "complete" causal network would refer to a single, integrated network, i.e., a network that could be represented as a connected graph. An incomplete causal net would refer to a graph in which there exists two nodes for which no path could be drawn. Theorists have yet to propose such an account, but it is a reasonable extension of existing proposals (see Ali et al., 2011). But, as we note above, a major constraint of the idea is that causal networks need to distinguish between causal chain structures, common-cause structures, common-effect structures, and diamond structures: all four structures can be represented as connected graphs. As Experiments 1 and 2 show, however, people privilege causal chains over other structures.

A more plausible idea comes from research on how reasoners compare one explanation to another. Lombrozo and colleagues argue that the fewer unexplained root causes an explanation has, the higher the quality of the explanation (Pacer & Lombrozo, 2017; Lombrozo & Vasilyeva, 2017). It may be that in our experiments, reasoners focused on root causes over the internal causal structures. In other words, one-model problems in Experiments 1, 2, and 4 had just one root cause, whereas multiple-model problems had more than one root cause. Yet, an account based on considering root causes alone could not explain why participants distinguished causal chains from other structures (Experiment 1), even when those structures were matched on their root causes (Experiment 2).

Other factors may affect how people evaluate the completeness of an explanation. For instance, the expectations of the person processing an explanation may dictate what constitutes a complete or an incomplete explanation. Contrast this explanation,

> Customer A: What is causing the engine to smoke?
> Mechanic: The engine must have overheated.

with this one:

> Customer B: What is causing the engine to smoke?
> Mechanic: The engine must have overheated, which caused the oil to be saturated with water. That forced the oil to overheat and flushed the cooling system with hot gases. The mixture of hot gases in the cooling system resulted in smoke.

The same mechanic provided two customers with explanations that differed in length. Which explanation is better? One might assume that an explanation's length is a cue to its fitness, because lengthy explanations provide a more detailed understanding of how explanandum came about. That is, the mechanic's second explanation contains many more causes and effects than the first, and so we might infer that the second conveys a better understanding of the physical mechanisms that result in engine smoke. However, an explanation's strength depends on context. On the one hand, the extra details may be useful and relevant for a customer with extensive knowledge and an interest in learning about what happened to their car – but for customers in a hurry, the explanation may seem unnecessarily verbose. On the other hand, customers with extensive background knowledge of what usually makes an engine smoke may prefer the first explanation. If a customer already has an understanding of, e.g., three different mechanisms that could lead to a car's engine smoking, then the mechanic's first explanation could narrow down the cause to one possible reason. In general, when asked for an explanation, reasoners might vary their responses by predicting the intentions and objectives of their audience, in particular their audience's preferred level of explanatory depth.

An explanation's scope may affect whether people perceive it as complete or incomplete (see, e.g., Johnson, Rajeev-Kumar, & Keil, 2016; Khemlani et al., 2011; Read & Marcus-Newhall, 1993; Sussman et al., 2014). Some explanations have broader scope than others. For instance, consider two explanations for why your friend Devon never showed up at a social event:

> Explanation 1: Devon had a scheduling conflict.
> Explanation 2: Devon is clinically depressed.

The first explanation has narrower scope than the second, e.g., it can explain Devon's absence from only one event, whereas the second explanation can account for absences from many events. It may be that reasoners construe explanations with broader scope as more incomplete and in need of further elaboration. Concerned friends may hypothesize about, e.g., when the depression started, what caused it, and the other things it could cause Devon to do, such as engage in substance abuse.

These, and other factors, could affect how people assess whether an explanation is complete or incomplete. At present, the model theory and the completeness hypothesis best account for how reasoners detect explanatory incompleteness in the first place. We conclude by considering how the detection of explanatory completeness can inform theories of how reasoners generate explanations.

6.3. The halting problem in generating explanations

Theorists have proposed several predictors of explanatory fitness, including: an explanation's simplicity (e.g., Walker et al., 2017), its coherence (e.g., Read & Marcus-Newhall, 1993), its relevance (e.g., Hilton, 1996), and its scope (Khemlani et al., 2011). Explanatory completeness may be yet another predictor of fitness. We argue, however, that completeness is unlike any of the extant predictors of fitness because the detection of completeness is fundamental to how reasoners generate explanations (Horne et al., 2019). When reasoners generate a novel explanation, the primary way they know not to generate antecedent causes ad infinitum is if they evaluate the completeness of their explanation. Suppose you're expecting a phone call from a friend, but you don't get a call. You might explain its absence by generating the following explanation:

> Your friend is unable to make the call.

You could accept the explanation as is, or you could elaborate on it by generating a preceding cause, e.g.:

> Your friend does not have reception, and is unable to make the call.

The failure to generate a complete explanation may cause you to

generate additional preceding causes, e.g., to consider what caused her to go to an area without reception, and so on. And those deliberations can result in delays of decisions, such as whether to call your friend or to wait longer. Incomplete explanations can therefore impose a cognitive and emotional tax, and they can deter reasoners from making decisions on the basis of their explanation. Yet as Miyake (1986) noted, people often reach stopping places in their generation of explanations – points in the process at which they deem an explanation sufficiently complete. The present theory provides a foundation for a more comprehensive account of when and how people choose to stop generating an explanation: a precondition for halting may be to infer explanatory completeness.

CRediT authorship contribution statement

Joanna Korman: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing - original draft. **Sangeet Khemlani:** Conceptualization, Data curation, Formal analysis, Funding acquisition, Methodology, Project administration, Resources, Supervision, Visualization, Writing - original draft.

Acknowledgements

We thank Paul Bello, Monica Bucciarelli, Ruth Byrne, Felipe de Brigard, Andrei Cimpian, Tony Harrison, Hillary Harner, Laura Hiatt, Zach Horne, Phil Johnson-Laird, Laura Kelly, Bertram Malle, Robert Mackiewicz, Greg Trafton, and Jeff Zemla for advice. We also thank Kalyan Gupta, Kevin Zish, and Knexus Research Corp. for their assistance in data collection.

Appendix A. Materials used in Experiments 1–4

Material	Domain	Description	Event 1	Event 2	Event 3	Event 4	Event 5
1	Biological	Hinolus are birds who live in the mountains of China.	…have slow digestion	…have thin feathers	…have Hexadrine ions in their blood	…have excellent night vision	…tolerate cold temperatures
2	Biological	Roobans are large animals living in Central Canada.	…eat mostly plants	…have pointed ears	…have thick coats	…hear high-frequency sounds	…have low levels of Meretin hormone
3	Mechanical	The Andon is a machine used alongside large construction equipment.	…engage a Tark actuator	…produce steam	…move its pistons	…pull a metal lever	…create a large shaft
4	Mechanical	The Zindo is a mechanical device used in factory assembly lines.	…release a valve	…engage a pump	…narrow an aperture	…open a glass pane	…flip a switch
5	Natural	Nerron Caves are found in the Eastern Hemisphere.	…have a limestone bed	…absorb natural acid	…form at cold temperatures	…become enlarged over time	…accumulate sediments
6	Natural	The Standinsk region is located in northern Europe.	…have trees with thick bark	…have abundant tall grass	…have mineral-rich soil	…have frequent heavy winds	…have crystalline rocks
7	Socioeconomic	The nation of Nelstadt is an island nation in the Pacific Ocean.	…create a new social welfare policy	…make gambling illegal	…sign an international treaty	…declare a new national holiday	…invest in scientific research
8	Socioeconomic	The Drenbow Corporation is located in Ireland.	…develop a new media strategy	…relocate its headquarters	…adjust the price of its products	…hire a new legal team	…remodel its stores

Appendix B. Schematics for complete and incomplete problems in Experiments 1–4 as well as examples of complete and incomplete problems using Material #4 in Appendix A above

Experiment	Completeness	Schematic	Example
1	Complete	*Event A causes Event B.*	Engaging a pump causes the Zindo to open a glass pane.
		Event B causes Event C.	Opening a glass pane causes the Zindo to narrow an aperture.
1	Incomplete (common-cause)	*Event A causes Event B.*	Engaging a pump causes the Zindo to open a glass pane.
		Event A causes Event C.	Engaging a pump causes the Zindo to narrow an aperture.
1	Incomplete (common-effect)	*Event A causes Event C.*	Engaging a pump causes the Zindo to open a glass pane.
		Event B causes Event C.	Narrowing an aperture causes the Zindo to open a glass pane.
2	Complete	*Event A causes Event B.*	Engaging a pump causes the Zindo to open a glass pane.
		Event B causes Event C.	Opening a glass pane causes the Zindo to narrow an aperture.
		Event C causes Event D.	Narrowing an aperture causes the Zindo to release a valve.
		Event D causes Event E.	Releasing a valve causes the Zindo to flip a switch.
2	Incomplete (diamond structure)	*Event A causes Event B.*	Engaging a pump causes the Zindo to open a glass pane.
		Event A causes Event C.	Engaging a pump causes the Zindo to narrow an aperture.
		Event C causes Event D.	Narrowing an aperture causes the Zindo to release a valve.
		Event B causes Event D.	Opening a glass pane causes the Zindo to release a valve.
3[a]	Complete	*Event A causes Event B.*	Engaging a pump causes the Zindo to open a glass pane.
		Event B causes Event C.	Opening a glass pane causes the Zindo to narrow an aperture.
3	Incomplete	*Event A causes Event X.*	Engaging a pump causes the Zindo to flip a switch.
		Event B causes Event C.	Opening a glass pane causes the Zindo to narrow an aperture.
4	Complete	*Event A causes Event B.*	Engaging a pump causes the Zindo to open a glass pane.
		Event B causes Event C.	Opening a glass pane causes the Zindo to narrow an aperture.
		Event C causes Event D.	Narrowing an aperture causes the Zindo to release a valve.
4[b]	Incomplete	*Event A causes Event X.*	Engaging a pump causes the Zindo to flip a switch.
		Event B causes Event C.	Opening a glass pane causes the Zindo to narrow an aperture.
		Event C causes Event D.	Narrowing an aperture causes the Zindo to release a valve.

[a] Experiment 3 contained both two- and three- premise problems. The example above shows only the two-premise case.
[b] Experiment 4 contained incomplete problems with a gap at both the *B* event (shown above) and at the *C* event (not shown).

References

Ali, N., Chater, N., & Oaksford, M. (2011). The mental representation of causal conditional reasoning: Mental models or causal models. *Cognition, 119*, 403–418.

Batygin, K., & Brown, M. E. (2016). Evidence for a distant giant planet in the solar system. *The Astronomical Journal, 151*, 22.

Frosch, C. A., & Johnson-Laird, P. N. (2011). Is everyday causation deterministic or probabilistic? *Acta Psychologica, 137*, 280–291.

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

Goodwin, G. P. (2014). Is the basic conditional probabilistic? *Journal of Experimental Psychology: General, 143*(3), 1214.

Goodwin, G. P., & Johnson-Laird, P. N. (2005). Reasoning about relations. *Psychological Review, 112*, 468–493.

Hall, N. (2000). Causation and the price of transitivity. *Journal of Philosophy, 97*, 198–222.

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

Hempel, C. (2002). Two models of scientific explanation. In Y. Balashov, & A. Rosenberg (Eds.). *Philosophy of science: Contemporary readings* (pp. 45–55). London: Routledge (Original work published 1965).

Hempel, C. G., & Oppenheim, P. (1948). Studies in the Logic of Explanation. *Philosophy of science, 15*(2), 135–175.

Hilton, D. (1996). Mental models and causal explanation: Judgements of probable cause and explanatory relevance. *Thinking and Reasoning, 2*, 273–308.

Horne, Z., Muradoglu, M., & Cimpian, A. (2019). Explanation as a cognitive process. *Trends in Cognitive Sciences, 23*, 187–199.

Johnson, S. G., Rajeev-Kumar, G., & Keil, F. C. (2016). Sense-making under ignorance. *Cognitive psychology, 89*, 39–70.

Johnson-Laird, P. N. (1983). *Mental models.* Cambridge: Cambridge University Press (Cambridge, MA: Harvard University Press).

Johnson-Laird, P. N., & Khemlani, S. (2017). Mental models and causation. In M. Waldmann (Ed.). *Oxford handbook of causal reasoning* (pp. 1–42). Elsevier, Inc.: Academic Press.

Johnson-Laird, P. N., Legrenzi, P., & Girotto, V. (2004). How we detect logical inconsistencies. *Current Directions in Psychological Science, 13*, 41–45.

Johnson, S. G., & Ahn, W. K. (2015). Causal networks or causal islands? The representation of mechanisms and the transitivity of causal judgment. *Cognitive science, 39*(7), 1468–1503.

Josephson, J. R. (2000). Smart inductive generalizations are abductions. *Abduction and induction* (pp. 31–44). Netherlands: Springer.

Keil, J. (2006). Explanation and understanding. *Annual Review of Psychology, 57*, 227–254.

Kelly, L. J., Khemlani, S., & Johnson-Laird, P. N. (2020). *Reasoning about duration.* (under review). (Manuscript under review).

Khemlani, S. (2018). Reasoning. In S. Thompson-Schill (Ed.). *Stevens' handbook of experimental psychology and cognitive neuroscience*. Wiley & Sons.

Khemlani, S., Barbey, A., & Johnson-Laird, P. N. (2014). Causal reasoning with mental models. *Frontiers in Human Neuroscience, 8*, 849.

Khemlani, S., Wasylyshyn, C., Briggs, G., & Bello, P. (2018). Mental models and omissive causation. *Memory & Cognition, 46*, 1344–1359.

Khemlani, S., Goodwin, G. P., & Johnson-Laird, P. N. (2015). Causal relations from kinematic simulations. In R. Dale, C. Jennings, P. Maglio, T. Matlock, D. Noelle, A. Warlaumont, & J. Yoshimi (Eds.). *Proceedings of the 37th annual conference of the cognitive science society* (pp. 1075–1080). Austin, TX: Cognitive Science Society.

Khemlani, S., & Johnson-Laird, P. N. (2011). The need to explain. *Quarterly Journal of Experimental Psychology, 64*, 2276–2288.

Khemlani, S., & Johnson-Laird, P. N. (2012). Hidden conflicts: Explanations make inconsistencies harder to detect. *Acta Psychologica, 139*, 486–491.

Khemlani, S., & Johnson-Laird, P. N. (2013). Cognitive changes from explanations. *Journal of Cognitive Psychology, 25*, 139–146.

Khemlani, S., & Johnson-Laird, P. N. (2017). Illusions in reasoning. *Minds & Machines, 27*, 11–35.

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

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*, 16766–16771.

Khemlani, S., Orenes, I., & Johnson-Laird, P. N. (2012). Negation: A theory of its meaning, representation, and use. *Journal of Cognitive Psychology, 24*, 541–559.

Khemlani, S., Sussman, A., & Oppenheimer, D. (2011). Harry Potter and the sorcerer's scope: Latent scope bias in explanatory reasoning. *Memory & Cognition, 39*, 527–535.

Khemlani, S., Wasylyshyn, C., Briggs, G., & Bello, P. (2018). Mental models and omissive causation. Manuscript in press at *Memory & Cognition*.

Lombrozo, T., & Vasilyeva, N. (2017). Causal explanation. *Oxford handbook of causal reasoning* (pp. 415–432).

McDermott, M. (1995). Redundant causation. *British Journal for the Philosophy of Science, 40*, 523–544.

Miyake, N. (1986). Constructive interaction and the iterative process of understanding. *Cognitive Science, 10*, 151–177.

Pacer, M., & Lombrozo, T. (2017). Ockham's razor cuts to the root: Simplicity in causal explanation. *Journal of Experimental Psychology: General, 146*, 1761–1780.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed.). NY: Cambridge University Press.

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

Railton, P. (1981). Probability, explanation, and information. *Synthese, 48*, 233–256.

Read, S. (1988). Conjunctive explanations: The effect of a comparison between a chosen and a nonchosen alternative. *Journal of Experimental Social Psychology, 24*, 146–162.

Read, S., & Marcus-Newhall, A. (1993). Explanatory coherence in social explanations: A parallel distributed processing account. *Journal of Personality and Social Psychology, 65*, 429–447.

Rehder, B., & Hastie, R. (2004). Category coherence and category-based property induction. *Cognition, 91*, 113–153.

Reichenbach, H. (1956). *The direction of time.* Berkeley: University of Los Angeles Press.

Rescher, N. (1995). *Satisfying reason: Studies in the theory of knowledge.* Dordrecht: Kluwer Academic Publishers.

Rips, L. (2010). Two causal theories of counterfactual conditions. *Cognitive Science, 34*, 175–221.

Salmon, W. (1978). Why ask, "Why?" An inquiry concerning scientific explanation. *Proceedings and Addresses of the American Philosophical Association, 51*, 683–705.

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

Sloman, S. A., Barbey, A. K., & Hotaling, J. (2009). A causal model theory of the meaning of "cause," "enable," and "prevent.". *Cognitive Science, 33*, 21–50.

Sussman, A., Khemlani, S., & Oppenheimer, D. (2014). Latent scope bias in categorization. *Journal of Experimental Social Psychology, 52*, 1–8.

Von Sydow, M., Hagmayer, Y., & Meder, B. (2016). Transitive reasoning distorts induction in causal chains. *Memory & Cognition, 44*(3), 469–487.

Walker, C. M., Bonawitz, E., & Lombrozo, T. (2017). Effects of explaining on children's preference for simpler hypotheses. *Psychonomic Bulletin & Review, 24*, 1538–1547.

Zemla, J. C., Sloman, S., Bechlivanidis, C., & Lagnado, D. A. (2017). Evaluating everyday explanations. *Psychonomic Bulletin and Review, 24*, 1488–1500.