Title

Inferring possible causes from event structures

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 48(0)

Author

Khemlani, Sangeet

Publication Date

2026

Inferring possible causes from event structures

Sangeet Khemlani
sangeet.s.khemlani.civ@us.navy.mil
U.S. Naval Research Laboratory, Washington, DC 20375 USA

Abstract

Causal reasoning is inherently temporal in nature; humans can infer whether one event is a possible cause of another, even from temporal descriptions. For example, the sentence, "after the book was placed, the shelf collapsed" implies that it's *possible* (but not necessary) that the book's placement caused the shelf to collapse, and that it's *impossible* that the shelf's collapse caused the book's placement. These modal inferences challenge existing accounts of causal reasoning. Two experiments reveal how humans infer possible causes from temporal descriptions. Experiment 1 showed that their possibility judgements depend on the event structures described by temporal relations. Experiment 2 varied the consistency of event descriptions and found that people inferred possible causes even from conflicting information. These data challenge extant theories of causal reasoning but corroborate the hypothesis that people comprehend causal relations by building mental simulations of event structures.

Keywords: causal relations; temporal relations; event structures; modal reasoning; mental simulation

Introduction

When humans reason about cause and effect, they must also reason about events and temporal concepts: time and causality are inextricably linked in human thinking. At least three sources of evidence substantiate this claim. First, the meanings of causal concepts denote temporal constraints. For instance, effects can't come before their causes. Hume (1748/1988) argued more strongly that causes *always* precede their effects, but Kant (1781/1934) noted that the two can take place at the same time, as when a flame causes candlewick to burn.

Second, causal relations, embodied in verbs such as *cause*, *allow*, and *prevent*, often take events as arguments. Consider the sentence, "The Dust Bowl caused mass migration," which treats *the Dust Bowl* and *mass migration* as two separate events that endured over long time periods. By definition, temporal relations, embodied in connectives such as *before*, *after*, and *during*, take events as arguments too. But events cannot be used as the arguments of spatial connectives, such as *next to*, which describe relations between two physical or conceptual objects, areas, or locations, as in: "California is next to Nevada." Spatial terms such as *next* can be co-opted to describe temporal descriptions of events, as in: "The next mass migration event will occur soon," but in this case, *next* is used as an adjective describing temporal sequence, and not a spatial relation.

Third, temporal information can often help establish causal relations between events. Temporal contiguity between events acts as a strong cue to a causal connection between them (Dickinson, 2001; Shanks, Pearson & Dickinson, 1989; Buehner & May, 2002, 2003) and that people estimate events that bear a causal connection to be closer together in time (Blakey et al., 2019; Buehner, 2012; Buehner & Humphreys, 2009; Haggard, Clark, & Kalogeras, 2002; Moore & Obhi, 2012). Lagnado and Sloman (2006) show how more abstract information about the temporal order between events, such as a computer virus first infecting a computer and then a computer crashing 20 minutes later, serves as a cue for reasoners to infer that the virus caused the crash (see also Bechlivanidis & Lagnado, 2016), and they show how such cues can be misleading. Temporal cues can help resolve ambiguous causal perceptions (Choi & Scholl, 2006), but they can also lead to causal "illusions" (Thorstad & Wolff, 2016; see also Matute et al., 2015). Developmental studies likewise reveal how children use temporal information to comprehend causal structures (Burns & McCormack, 2009; Rottman, Kominsky, & Keil, 2014). Temporal and causal reasoning appears to develop along parallel trajectories (Lohse, Kalitschke, Ruthmann, & Rakoczy, 2015; McCormack, 2015; see also Zhang, Zhang, Gao, & Chen, 2026). And one imaging investigation designed to locate neural substrates for causal, temporal, and spatial processing showed greater overlap between causal and temporal processing than causal and spatial processing (Kranjec, Cardillo, Lehet, & Chatterjee, 2012).

Reasoners are imminently capable of drawing causal hypotheses from limited verbal information. For example, this sentence describes two events in a temporal order:

1. David put a book on the shelf.
   Then the shelf collapsed.

and reasoners infer a causal connection between David's actions and the shelf collapsing (Khemlani & Johnson-Laird, 2020), despite the absence of any specific causal verb, such as *cause* or *make.* In essence, they made causal bridging inferences (Clark, 1975) to make sense of the sequence of events. But few studies have investigated how people reason from temporal information to build causal hypotheses, despite abundant evidence of how causal reasoning and temporal reasoning interact. Consider this brief description of an improvised jam session between unacquainted musicians:

2. The guitarist began playing a melody.
   Then the drummer added a swing beat.
   Then the bassist added a walking bass line.
   Then the saxophonist added a progression.

With rudimentary knowledge of what these different instruments sound like, even musically untrained readers may be able to imagine the session. And nearly all capable adults should be able to make sensible inferences about the following questions:

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

3a. What could have caused the bassist to add their bass line?
b. What could not have caused the bassist's decision?

There exist many correct answers to (3a): the guitarist's melody, the drummer's rhythm, the bassist's own whims, or something else entirely might explain the bassist's decision. But since the sax began playing after the bassist, it had no impact on the bassist's decision, which serves as the only correct answer to (3b). Though the sequence of events in (2) are minimal, they permit a complex set of possible and impossible causal relations. All of the following are sensible deductions:

4a. The guitarist possibly caused the drummer's beat.
b. The guitarist possibly caused the bassist's bass line.
c. The guitarist possibly caused the saxophonist's melody.
d. The drummer couldn't have caused the guitar melody.
e. The bassist couldn't have caused the drummer's beat.
f. The sax couldn't have caused the bassist's bass line.

and so on. The answers in (4a-f) are "modal" in nature, i.e., they do not concern any definitive causal relation between events, but rather the possibility (or impossibility) of a causal relation.

This paper investigates how reasoners infer modal causes from event structures such as the one embodied in (2) above. It begins by reviewing how modal reasoning operates in human thinkers, and how it diverges from formal systems of reasoning, such as the infinite set of modal logics developed to explain many sensible deductive inferences. It also explains how the inference of possible causes from event structures serves to challenge contemporary cognitive theories of causation. And it reports two experiments that show how human reasoning about possible causes is both sensitive to event structures (Experiment 1) and resilient to inconsistent information (Experiment 2). It concludes by appraising existing accounts of human causal reasoning in light of the data.

Reasoning about possibilities and possible causes

Recent research reveals the central role possibilities play across many higher-level thought processes, including explanatory reasoning, moral decision-making, mechanical reasoning, and counterfactual thinking, on the mental representation of possibilities (Carey, Leahy, Redshaw, & Suddendorf, 2020; Johnson-Laird, Byrne, & Khemlani, 2024; Johnson-Laird & Ragni, 2019, 2024; Grigoroglou & Ganea, 2022; Phillips, Morris, & Cushman, 2019). Possibilities serve as the basis of both explanations and counterfactuals: explanations are possible ways of accounting for some phenomenon (Korman & Khemlani, 2020), and counterfactuals are situations that could have possibly happened but didn't (Byrne, 2016). As De Brigard and colleagues (2020) argue, episodic memories are based in the representation of possibilities as well, because possibilities can be rehearsed to improve future outcomes. Possibilities serve as the most rudimentary form of expressing and thinking about uncertainty, particularly compared to expressions of probability. Consider the examples in (5) below:

5a. It's possible that the elections were rigged.
b. It's probable that the elections were rigged.

The statement in (5a) conveys uncertainty without committing to any likelihood, whereas (5b) describes a high probability, depending on the context, but certainly one that is > 50% (Wallsten, Fillenbaum, & Cox, 1986; McGlone & Reed, 1998; and see Lassiter, 2017 for broader exposition on expressions of uncertainty). Hence (5a) expresses uncertainty without probability. In line with the primacy of possibilities in human thinking, Gauffroy and Barrouillet (2011) show that reasoning about possibilities develops earlier than reasoning about truth and falsity.

If, as previous research suggests, possibilities are a bedrock representation that subserves human thinking and reasoning capabilities, as previous research suggests, then much of language should bear a "modal" semantics, i.e., meanings that reduce to sets of possible states of the world (see, e.g., Johnson-Laird et al., 2024; Khemlani, Byrne, & Johnson-Laird, 2018). As a consequence, reasoners should have no difficulty drawing modal conclusions from non-modal expressions. Hinterecker and colleagues (2016) ran studies that examined such patterns of inference for everyday arguments, e.g.,

6. The U.S. will ratify the Kyoto protocol and commit to reducing CO2 emissions, or global temperatures will reach a theoretical point of no return in the next 100 years, or both. Therefore, possibly, the U.S. will ratify the Kyoto protocol.

A schematic for the argument in (6) is as follows:

7. A or B or both.
Therefore, possibly A.

where A = CO2 reductions and B = rising global temperatures. Normal modal logics treat such inferences as invalid (see Khemlani et al., 2018; but cf. Bringsjord & Govindarajulu, 2020), but reasoners accepted the conclusion (*possibly A*) on 91% of the trials. Perhaps a more subtle result is that non-modal expressions, such as disjunctions of the form, *A or B or both*, can yield modal conclusions – a pattern that normal modal logics treat as invalid for all but the most trivial arguments (e.g., *A; therefore possibly A*, is valid in modal logic S5; see Hughes & Cresswell, 1996). Indeed, Khemlani et al. (2018) extend this prediction to all connectives in natural language, including conditionals, conjunctions, and negations – the meanings of *if*, *and*, and *not* should all concern possibilities.

The primacy of possibilities in human thinking can explain how humans infer possible causes from temporal expressions. Consider the following statement, which integrates the first two sentences in example (2) above:

8. The guitarist began playing a melody before the drummer added a swing beat.

In (8), the temporal connective *before* links two events (the guitar melody and the swing beat). Temporal connectives such as *before*, *after, during*, and *while* help reasoners construct complex event structures to keep track of when events occur in relation to one another. Both logicians and cognitive scientists have explored their semantics: logicians developed temporal logics to describe how to deduce valid conclusions from statements such as (8). For example, (8) implies that the swing beat started *after* the guitar melody began, and any viable formal calculus, such as Allen's interval algebra (1983), should make it possible to prove *B happened after A* from *A happened before B*. Cognitive scientists have argued that people simulate possibilities to represent event structures (e.g., Kelly & Khemlani, 2023a; Khemlani, Harrison, & Trafton, 2015; Schaeken, Johnson-Laird, & d'Ydewalle, 1996; Ulrich & Maienborn, 2010; Vandierendonck, 1996), and more generally that these event structures are encoded spatially, akin to a mental timeline (Knauff, 2013; Núñez & Cooperrider, 2013; Tversky, 2019). But perhaps previous studies of temporal reasoning were limited in their scope. They focused on how people draw definitive, necessary temporal conclusions based on temporal premises such as (8). As the previous section demonstrates, however, reasoners can draw modal conclusions even from non-modal expressions, such as compound sentences linked by the sentential connectives *if, and*, and *or*. Hence, sentences that describe the structures of events may yield sensible modal conclusions too.

What sorts of modal conclusions can (8) yield? Consider again the intimate conceptual link between temporal and causal relations: causes and their effects are often construed as events ordered in time. Recent theorists concur that many causal verbs, such as *cause* and *enable*, refer to what is possible and what is impossible, or what was counterfactually possible and impossible – hence, they, too, bear a modal semantics (Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2021; Goldvarg & Johnson-Laird, 2001; Khemlani, Bello, Briggs, Harner, & Wasylyshyn, 2021; Wolff, 2007). Taken together, statements such as (8) may yield sensible causal conclusions, such as:

9a. It's possible that the guitarist's melody caused the drummer's swing beat.
 b. It's not possible that the drummer's swing beat caused the guitarist's melody.

Such conclusions may have been overlooked by previous investigations because they are modal in nature: they concern, not causal facts but rather a causal possibility (9a) and a causal impossibility (9b).

The remainder of the paper reports two experiments that evaluated the ability to draw modal conclusions about causality from temporal premises. Both presented participants with problems that described a sequence of actions carried out by a secret agent, e.g.,

10. Suppose that an intelligence report about Agent X reveals the following information:
    X will pick the lock before they talk to the lawyer. X will shred the documents while they talk to the lawyer. X will configure the satellite hookup after they shred the documents.

Participants in the studies were asked questions of the following form:

11. Is it possible that configuring the satellite hookup caused Agent X to pick the lock?

The premises in (10) imply – but do not state explicitly – that Agent X will configure the satellite after they pick the lock. Hence, it should be *impossible* that configuring the satellite would cause the Agent to pick the lock, since doing so would violate the constraint that effects never precede their causes. Hence (11) is intuitively invalid – and yet no existing logical calculus explains its validity: such a system would need to combine a modal logic with temporal and causal axioms. Experiment 1 tested whether reasoners would endorse such modal conclusions, and more generally, whether their endorsements were sensitive to the event structure described by the premises. Experiment 2 presented participants with logically inconsistent structures to test whether they drew modal conclusions despite the conflicts. Both experiments revealed capable and flexible reasoning behaviors.

Experiment 1

Experiment 1 sought to test whether event structures guide endorsements of modal causal conclusions. Participants read problems that described the actions of a secret agent using temporal language, and they assessed whether a particular causal relation between those actions was possible or not. The experiment manipulated structural and surface level properties of the problems along several dimensions, the most pertinent of which being whether the question described a potential causal event that happened before or after its effect: half of the questions described causes that occurred before events, and half described the opposite, as in (11) above. The study also varied whether the vignette described events in the past or the future – (11) concerns future events – and whether or not there was an intervening event between the two described in the question, as in (11).

Method

Participants. 60 participants (mean age = 41.6 years; 20 females, 39 males, 1 other) were recruited through Amazon Mechanical Turk for monetary compensation. All participants spoke English as a first language.

Materials and task. Each problem described an intelligence report that described a secret agent's actions. Reports consisted of three sentences that each described a temporal relation (*before*, *after*, or *while*) between two of the agent's actions, e.g., "[Agent] picked the lock before they talked to a lawyer." Problems concerned 4 separate actions. The experiment assigned actions to each problem by randomly sampling from a pool of 15 espionage-relevant actions that were semantically unrelated to one another. Participants assessed whether a causal relation was possible by answering questions of the form: "Is it possible that [action 1] caused X to [action 2]?" They responded by selecting one of three buttons marked "yes, it's possible", "no, it's impossible", and "there's not enough information to be certain".

Design and procedure. Half the problems concerned questions about causes that preceded effects, and half concerned effects that preceded causes. Half the problems concerned past events (using simple past tense, e.g., "picked the lock") and the other half concerned future events (e.g., "will pick the lock"). And half the problems concerned questions that described events that were chronologically adjacent to one another, and while the other half described events that were separated by an intervening event. The experiment therefore implemented a 2 (question structure: cause before effect vs. cause affect effect) × 2 (past vs. future events) × 2 (number of intervening events between cause and effect: 0 vs. 1) repeated measures design. Participants carried out 16 separate problems that included the 8 problem schematics (see Appendix A) describing events in the past or future. The experiment randomly assigned the order of the problems, the materials in each problem, and the order of the response buttons. Participants could take as long as they wished on each problem.

Open science. Data, materials, experimental code, and analyses are provided at https://osf.io/q8usr.

Results

Participants in Experiment 1 accepted modal conclusions about possible causes systematically. They considered putative causes that occurred before effects to be "possible" on 70% of trials vs. 22% when those causes came after their effects (Wilcoxon test, $z = 6.30$, $p < .0001$, Cliff's $\delta = .83$). Qualitative responses in the study were coerced to a numerical scale to create a composite measure of modal conclusion acceptance, such that "possible" = 1, "there's not enough info…" = 0.5, and "impossible" = 0.0 to create Figure 1, which depicts distributions of participants' responses as a function of the order of events in the question.

Other manipulations of the problem structure in the experiment had no effect on endorsements, i.e., there were no reliable difference in endorsement patterns for past and future events (45% "possible" responses for past events vs. 47% for future events; Wilcoxon test, $z = .71$, $p = .48$, Cliff's $\delta = .06$). Likewise, the number of events between the putative cause and its effect yielded no significant difference (47% "possible" for 0 intervening events vs. 45% for 1 intervening event; Wilcoxon test, $z = .28$, $p = .78$, Cliff's $\delta = .04$).

Experiment 1 showed that reasoners considered an event to be a possible cause when it came before its effect. None of the other manipulations in the study impacted performance. The results suggest that participants' tendency to rate a causal relation as "possible" depended on the event structure to which the temporal relations referred: temporal precedence was all that was needed to infer that an event could potentially cause another event.

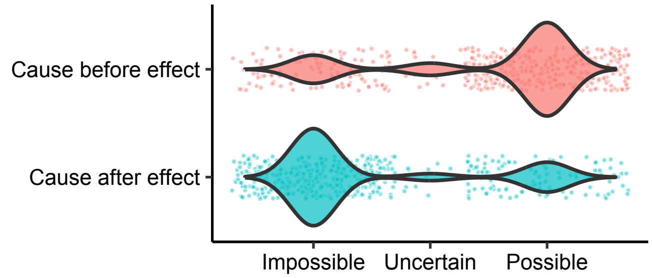

Figure 1. Violin density plots of the distributions of whether participants considered putative causes to be possible, impossible, or uncertain in Experiment 1, as a function of whether those causal events came before or after their effects.

The problems in Experiment 1 used sets of premises that were consistent, e.g., the premises in (10) could all be true at the same time. A hallmark of human reasoning, however, is how it copes with inconsistency: humans often do so by generating explanations that help them systematically reject or weaken the claims of the premises (Kelly & Khemlani, 2023b; Khemlani & Johnson-Laird, 2011, 2012, 2020). Inconsistent event structures may make it more likely for reasoners to judge an event as a possible cause of another – a result that contravenes existing theories of causation (see General Discussion). Experiment 2 provided participants with inconsistent temporal premises to test this hypothesis.

Experiment 2

The previous experiment showed that reasoners base modal causal inferences on event structures inferred by the premises. What happens when those structures break down in face of inconsistent information? Consider this problem:

12. Suppose that an intelligence report about Agent X reveals the following information:

    X hacked the smartphone before they reviewed surveillance. X configured the satellite hookup after they reviewed surveillance. X decrypted the hard drive while they hacked the smartphone. ***X configured the satellite hookup before they hacked the smartphone.***

    Is it possible that configuring the satellite hookup caused X to decrypt the hard drive?

The premises in (12) are inconsistent: the last premise (in **bold** above for emphasis) cannot be true if the rest of the premises are true, because the first two premises imply that X hacked the smartphone before they configured the satellite hookup. Given this conflict, how might reasoners respond to the question? If such conflicts introduce uncertainty, reasoners may prefer to respond that "there's not enough information to be certain." Alternatively, the first three premises in (12) imply that X configured the satellite hookup after decrypting the hard drive. Hence, reasoners who satisfice (i.e., use only as much information as is necessary to respond to the question) should respond "no". But, reasoners who process all of the information may process the inconsistency *locally*, in a flexible manner that permits them to consider the consequences of what would happen if the last premise was true. One consequence is that configuring the satellite hookup would happen before X decrypted the hard drive. In this case, reasoners who recognize this possibility should respond “yes.” Experiment 2 made use of problems akin to (12) to test these different patterns of response.

Method

Participants. 60 participants were recruited through Amazon Mechanical Turk for monetary compensation, but 4 were rejected for not providing sensible English responses in post-experimental questionnaires, and so the remaining 56 participants were subjected to analysis (mean age = 40.5 years; 23 females, 31 males, 1 non-binary, 1 prefer not to say). All participants spoke English as a first language.

Materials, design, task, and procedure. Materials came from the same pool of items used in Experiment 1. Each problem consisted of four sentences that described temporal relations (*before*, *after*, or *while*) concerning a secret agent’s actions. Problems concerned 4 separate actions – but unlike Experiment 1, half the problems in Experiment 2 were inconsistent. Inconsistencies were introduced such that the 4th premise could not be possible if the previous three were true, as in (12). The experiment varied whether questions concerned causes that came before vs. after their effects, and whether the premises concerned events in the past or events in the future. The experiment implemented a 2 (question structure: cause before effect vs. cause affect effect) × 2 (past vs. future events) × 2 (consistent vs. inconsistent 4th premise) repeated measures design. As in Experiment 1, participants assessed whether a causal relation was possible by answering questions of the form: “Is it possible that [action 1] caused X to [action 2]?” They responded by selecting one of three buttons marked “yes, it’s possible”, “no, it’s impossible”, and “there’s not enough information to be certain”. Participants carried out 8 separate problems that included problem schematics (see Appendix B). The experiment employed similar randomization schemes as the previous study, i.e., it randomized the order of the problems, the materials, and the order the response buttons. Participants could take as long as they wished on each problem.

Open science. Data, materials, experimental code, and analyses are provided at https://osf.io/q8usr.

Results

Experiment 2 replicated the results of Experiment 1 for consistent problems: participants accepted modal conclusions about possible causes when causal events came before their effects (66% “possible” responses vs. 47% when causal events occurred after their effects; Wilcoxon test, z = 4.21, p < .0001; Cliff’s δ = .41). And as in previous study, the effect held regardless of whether the events were in the past

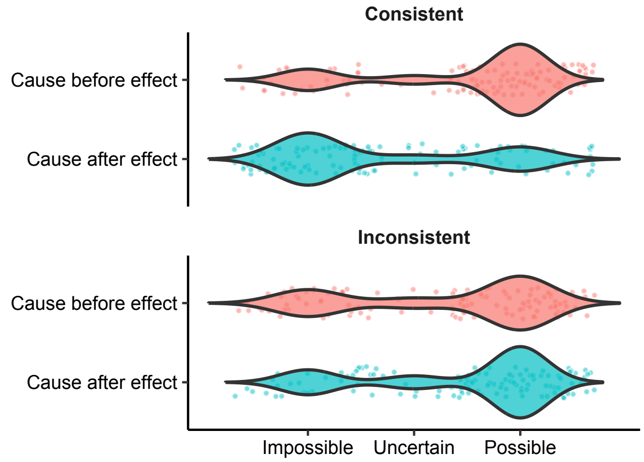

Figure 2. Violin density plots of the distributions of whether participants considered putative causes to be possible, impossible, or uncertain in Experiment 2, as a function of whether those causal events came before or after their effects, and as a function of whether the premises were consistent or inconsistent.

(56% “possible” responses) or the future (57% “possible” responses; Wilcoxon test, z = .42, p = .67; Cliff’s δ = .04). Experiment 2 also revealed a significant effect of consistency: participants responded “yes, it’s possible” more often for inconsistent than consistent problems (63% vs. 50%; Other manipulations of the problem structures in the experiment had no effect on endorsements, i.e., there were no reliable difference in endorsement patterns for past and future events (45% “possible” responses for past events vs. 47% for future events; Wilcoxon test, z = 3.27, p = .001, Cliff’s δ = .27). Figure 2, which plots the same composite measure of modal conclusion acceptance as in Figure 1, reveals the disparity between consistent and inconsistent problems: participants tended to respond “yes, it’s possible” for causal events that came after their effects. Hence, the interaction between question structure and inconsistency was reliable (Wilcoxon test, z = 5.15, p < .0001, Cliff’s δ = .56).

These results suggest significant flexibility when reasoning about inconsistencies, and reveal a previously unknown strategy for conflict resolution: the conflict prompted reasoners consider the possibility that a putative causal event that should’ve otherwise come after its effect (had the premises been consistent) may have come before it – and as a result, it permitted them to draw a sensible modal conclusion.

General discussion

Two experiments revealed that people can reason from temporally ordered events to possible causes. Experiment 1 showed that their tendency to rate one event as a “possible cause” of another was driven by temporal precedence: putative causes that came before their effects were considered possible. Experiment 2 showed that this effect is modulated by whether the premises were consistent or not: inconsistent premises allowed reasoners to consider possibilities that they otherwise wouldn’t have, and as a result, they assessed a cause as possible when it was consistent with at least some subset of the premises. The experiment showed how conflicts don't flummox reasoners or introduce uncertainty: rather, they served to bring to mind alternative possibilities.

Studies of causal reasoning have overlooked how people make inferences about causal possibilities. The results reported here appear to challenge many existing frameworks and theories that account for causal reasoning. The rest of the discussion summarizes the central hypotheses that form the foundations of many of these accounts, and marshals the evidence in light of them.

Graded causation. Many theories of causation treat the concept as a continuous, graded measure that denotes either the strength of a causal relation (e.g., Stephan, Engelmann, & Waldmann, 2023), causal power (Cheng, 1997), a posterior in a causal Bayes net (e.g., Sloman & Lagnado, 2015), or as a causal "force" that is transmitted from an actor to a patient (e.g., Wolff, 2007). Each of these views considers a cause to be inherently graded (but cf. O'Neill, Henne, Bello, Pearson, & De Brigard, 2022 for a critical evaluation). Graded notions of causality cannot explain the bimodal distributions of responses in Experiments 1 and 2. More generally, they cannot explain why reasoners should consider any potential event as "impossible": an "impossible" event corresponds to an event with a probability of 0 – but even an event with a miniscule probability, e.g., P(decrypted hard drive | satellite hookup) = .00001 is technically "possible". If, each of the premises in (10) above are treated as uncertain and graded, there should be in principle and in practice no way of arriving at a probability of exactly 0.

Another difficulty with these theories is how they might treat temporally coded information. Many graded accounts of causal reasoning assume that the information to be processed is causal in nature (e.g., Wolff, 2007), but such theories make no claims or commitments about the representation of temporal relations, and therefore have no mechanisms to explain how modal causes can be inferred from event structures.

Causation as physical and counterfactual simulation. Recent computational theories of causal reasoning argue that reasoning about causality is akin to building a counterfactual simulation of what would have happened if a putative causal factor hadn't been present (Gerstenberg, 2024; Gerstenberg et al., 2021). The simulations these authors have explored are physical simulations computed using a physics engine. One advantage to using a physics engine is that, by their nature, they are capable of tracking metric time and in principle have a built-in representation for integrating temporal information. No counterfactual simulation theories have availed themselves of these capacities as far as is known, but at least in principle they can be extended to account for how people represent temporal information.

But, if causality was a concept that could be reduced to counterfactual considerations, as these theorists argue, then answers should have been modulated by whether events were framed to concern future events, because counterfactuals by definition concern past, not future events. Only past events can be contrasted against real world facts, and so counterfactual theorists should predict that it's easier to infer possible causes for past events than future ones. Yet neither experiment uncovered such patterns. One additional challenge to such systems is in how they could process inconsistent information, as in (12) above. Just as physical situations in the real world are never inconsistent, physics engines are unable to represent or cope with inconsistencies, and so novel mechanisms for handling conflicts must be introduced to reason about them.

Mental models of causation. One treatment of causation holds that people draw causal inferences by constructing small scale, temporally ordered possibilities – mental models – of a cause and its effect (Goldvarg & Johnson-Laird, 2001; Khemlani, Barbey, & Johnson-Laird, 2014). It argues that to represent, e.g., *A causes B*, reasoners consider the possibility in which *A occurs and then B occurs*, as well as alternative possibilities, such as one where *A doesn't occur and B doesn't occur*. The theory argues that because people represent possibilities, all reasoning is fundamentally modal in nature. And it shows how causal mental models are derived from temporal information, and it likewise shows how reasoners can detect inconsistencies in temporal information (Kelly, Khemlani, & Johnson-Laird, 2020). Hence, the theory readily handles many of the results in Experiments 1 and 2: it explains how temporal information can give rise to modal causal conclusions. It likewise explains why past and future events made no difference on performance, since reasoners can build mental models of causes in the past and the future.

One limiting factor of the theory, however, is its treatment of what individuals do when they arrive at an inconsistency. Previous computational models and theoretical accounts argued that reasoners detect inconsistency when they fail to build a model of a situation, or in other words, construct a "null" model that bears no content. But Experiment 2 suggests that reasoners employed an alternative strategy: they may have considered multiple models of different consistent subsets of the premises. If such a strategy is indeed robust, then model-based theories should be updated to incorporate and provide computational accounts of the strategy

In sum, reasoners are adept at inferring hypothetical causes of some set of outcomes, even when the only information available has no causal content. Hume (1748/1988) once argued that all forms of reasoning about facts were "founded on the relation of cause and effect." These experiments are the first to explore how reasoners infer possible causes from the structures of events, and reveal new insights into how causal reasoning can be efficient, flexible, and strategic.

Acknowledgements

This work was supported a grant from the Naval Research Lab, and data were collected by Knexus Research Corporation. The research benefitted from feedback and critiques from Bill Adams, David Beltrán, Branden Bio, Gordon Briggs, Ruth Byrne, Tony Harrison, Phil Johnson-Laird, Markus Knauff, Maria Kon, Andrew Lovett, Marco Ragni, and Greg Trafton.

References

Allen, J. F. (1983). Maintaining knowledge about temporal intervals. *Communications of the ACM*, 26, 832-843.
Bechlivanidis, C. & Lagnado, D. A. (2016). Time reordered: Causal perception guides the interpretation of temporal order, *Cognition*, 146, 58-66.
Blakey, E., Tecwyn, E. C., McCormack, T., Lagnado, D. A., Hoerl, C., Lorimer, S., & Buehner, M. J. (2019). When causality shapes the experience of time: Evidence for temporal binding in young children. *Developmental Science*, 22, e12769.
Bringsjord, S., & Sundar Govindarajulu, N. (2020). Rectifying the mischaracterization of logic by mental model theorists. *Cognitive Science*, 44, e12898.
Buehner, M. J. (2012). Understanding the past, predicting the future: Causation, not intentional action, is the root of temporal binding. *Psychological Science*, 23, 1490-1497.
Buehner, M. J., & Humphreys, G. R. (2009). Causal binding of actions to their effects. *Psychological Science*, 20, 1221-1228.
Buehner, M. J., & May, J. (2002). Knowledge mediates the timeframe of covariation assessment in human causal induction. *Thinking & Reasoning*, 8, 269-295.
Buehner, M. J., & May, J. (2003). Rethinking temporal contiguity and the judgement of causality: Effects of prior knowledge, experience, and reinforcement procedure. *Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 56, 865-890.
Burns, P., & McCormack, T. (2009). Temporal information and children's and adults' causal inferences. *Thinking & Reasoning*, 15, 167-196.
Byrne, R. M. (2016). Counterfactual thought. *Annual Review of Psychology*, 67, 135-157.
Carey, S., Leahy, B., Redshaw, J., & Suddendorf, T. (2020). Could it be so? the cognitive science of possibility. *Trends in Cognitive Sciences*, 24, 3-4.
Cheng, P. W. (1997). From covariation to causation: A causal power theory. *Psychological Review*, 104, 367.
Choi, H., & Scholl, B. J. (2006). Perceiving causality after the fact: Postdiction in the temporal dynamics of causal perception. *Perception*, 35, 385-399.
Clark, H. H. (1975). Bridging. In *Theoretical Issues in Natural Language Processing*.
De Brigard, F., Gessell, B., Yang, B. W., Stewart, G., & Marsh, E. J. (2020). Remembering possible times: Memory for details of past, future, and counterfactual simulations. *Psychology of Consciousness: Theory, Research, and Practice*, 7, 331.
Dickinson, A. (2001). Causal learning: An associative analysis. *Quarterly Journal of Experimental Psychology*, 49, 60-80.
Gauffroy, C., & Barrouillet, P. (2011). The primacy of thinking about possibilities in the development of reasoning. *Developmental Psychology*, 47, 1000.
Gerstenberg, T. (2024). Counterfactual simulation in causal cognition. *Trends in Cognitive Sciences*, 28, 924-936.
Gerstenberg, T., Goodman, N. D., Lagnado, D. A., & Tenenbaum, J. B. (2021). A counterfactual simulation model of causal judgments for physical events. *Psychological Review*, 128, 936.
Goldvarg, E., & Johnson-Laird, P. N. (2001). Naive causality: A mental model theory of causal meaning and reasoning. Cognitive Science, 25, 565-610.
Grigoroglou, M., & Ganea, P. A. (2022). Language as a mechanism for reasoning about possibilities. *Philosophical Transactions of the Royal Society B*, 377, 20210334.
Haggard, P., Clark, S., & Kalogeras, J. (2002). Voluntary action and conscious awareness. *Nature Neuroscience*, 5, 382-385.
Hegarty, M. (2004). Mechanical reasoning by mental simulation. *Trends in Cognitive Sciences*, 8, 280-285.
Hughes, G. E., & Cresswell, M. J. (1996). *A new introduction to modal logic.* Psychology Press.
Hume, D. (1748/1988). *An enquiry concerning human understanding.* (A. Flew, Ed.) La Salle, IL: Open Court.
Johnson-Laird, P.N., Byrne, R.M.J., & Khemlani, S. (2024). Models of possibilities instead of logic as the basis of human reasoning. *Minds and Machines*, 34, 19.
Johnson-Laird, P. N., & Ragni, M. (2019). Possibilities as the foundation of reasoning. *Cognition*, 193, 103950.
Johnson-Laird, P. N., & Ragni, M. (2024). Reasoning about possibilities: Modal logics, possible worlds, and mental models. *Psychonomic Bulletin & Review*, 1-28.
Kant, I. (1781/1934). *Critique of pure reason.* New York: Dutton.
Kelly, L., & Khemlani, S. (2023a). Iconicity bias and duration. *Journal of Experimental Psychology: Language, Memory, and Cognition*, 49, 1715-1731.
Kelly, L., & Khemlani, S. (2023b). Temporal explanations. *Journal of Experimental Psychology: General*, 152, 1639-1646.
Kelly, L. J., Khemlani, S., & Johnson-Laird, P. N. (2020). Reasoning about durations. *Journal of Cognitive Neuroscience*, 32, 2103-2116.
Khemlani, S. S., Barbey, A. K., & Johnson-Laird, P. N. (2014). Causal reasoning with mental models. *Frontiers in Human Neuroscience*, 8, 849.
Khemlani, S., Bello, P., Briggs, G., Harner, H., & Wasylyshyn, C. (2021). Much ado about nothing: The mental representation of omissive relations. *Frontiers in Psychology*, 11, 609658.
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., Harrison, A. M., & Trafton, J. G. (2015). Episodes, events, and models. *Frontiers in Human Neuroscience*, 9, 590.
Khemlani, S. & Johnson-Laird, P.N. (2011). The need to explain. *Quarterly Journal of Experimental Psychology*, 64, 2276-88.
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. (2020). Causal conflicts produce domino effects. *Quarterly Journal of Experimental Psychology*, 73, 2317-2327.
Korman, J., & Khemlani, S. (2020). Explanatory completeness. *Acta Psychologica,* 209, 103139.
Knauff, M. (2013). *Space to reason: A spatial theory of human thought.* MIT Press.
Kranjec, A., Cardillo, E. R., Schmidt, G. L., Lehet, M., & Chatterjee, A. (2012). Deconstructing events: the neural bases for space, time, and causality. *Journal of Cognitive Neuroscience*, 24, 1-16.
Lagnado, D. A., & Sloman, S. A. (2006). Time as a guide to cause. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32, 451.
Lassiter, D. (2017). *Graded modality: Qualitative and quantitative perspectives.* Oxford University Press.
Lohse, K., Kalitschke, T., Ruthmann, K., & Rakoczy, H. (2015). The development of reasoning about the temporal and causal relations among past, present, and future events. *Journal of Experimental Child Psychology*, 138, 54-70.
Matute, H., Blanco, F., Yarritu, I., Díaz-Lago, M., Vadillo, M. A., & Barberia, I. (2015). Illusions of causality: How they bias our everyday thinking and how they could be reduced. *Frontiers in Psychology*, 6, 888.
McCormack, T. (2015). The development of temporal cognition. *Handbook of Child Psychology and Developmental Science*, 2, 624-70.
McGlone, M. S., & Reed, A. B. (1998). Anchoring in the interpretation of probability expressions. *Journal of Pragmatics*, 30, 723-733.

Moore, J. W., & Obhi, S. S. (2012). Intentional binding and the sense of agency: A review. *Consciousness and Cognition*, 21, 546-561.
Núñez, R., & Cooperrider, K. (2013). The tangle of space and time in human cognition. *Trends in Cognitive Sciences*, 17, 220-229.
O'Neill, K., Henne, P., Bello, P., Pearson, J., & De Brigard, F. (2022). Confidence and gradation in causal judgment. *Cognition*, 223, 105036.
Rottman, B. M., Kominsky, J. F., & Keil, F. C. (2014). Children use temporal cues to learn causal directionality. *Cognitive Science*, 38, 489-513.
Schaeken, W., Johnson-Laird, P. N., & d'Ydewalle, G. (1996). Mental models and temporal reasoning. *Cognition*, 60, 205-234.
Shanks, D. R., Pearson, S. M., & Dickinson, A. (1989). Temporal contiguity and the judgement of causality by human subjects. The *Quarterly Journal of Experimental Psychology B: Comparative and Physiological Psychology*, 41, 139-159.
Sloman, S. A., & Lagnado, D. (2015). Causality in thought. *Annual Review of Psychology*, 66, 223-247.
Stephan, S., Engelmann, N., & Waldmann, M. R. (2023). The perceived dilution of causal strength. *Cognitive Psychology*, 140.
Thorstad, R., & Wolff, P. (2016). What causal illusions might tell us about the identification of causes. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 38).
Tversky, B. (2019). *Mind in motion: How action shapes thought.* Hachette: UK.
Ulrich, R., & Maienborn, C. (2010). Left–right coding of past and future in language: The mental timeline during sentence processing. *Cognition*, 117, 126-138.
Vandierendonck, A. (1996). Evidence for mental-model-based reasoning: A comparison of reasoning with time and space concepts. *Thinking & Reasoning*, 2, 249-272.
Wallsten, T. S., Fillenbaum, S., & Cox, J. A. (1986). Base rate effects on the interpretations of probability and frequency expressions. *Journal of Memory and Language*, 25, 571-587.
Wolff, P. (2007). Representing causation. *Journal of Experimental Psychology: General*, 136, 82.
Zhang, Y., Zhang, J., Gao, C., & Chen, Y. (2026). Asymmetry of temporal causal reasoning in preschoolers: The role of temporal direction representation. *Journal of Experimental Child Psychology*, 264, 106447.

Appendix A. Problem schematics used in Experiment 1.

Premise 1	Premise 2	Premise 3	Question	Question structure	# of intervening events
A before B	C after B	D while C	Is it possible that D caused B?	Cause after effect	0
A before B	C after B	D while C	Is it possible that B caused D?	Cause before effect	0
A before B	C after B	D while C	Is it possible that D caused A?	Cause after effect	1
A before B	C after B	D while C	Is it possible that A caused D?	Cause before effect	1
A before B	D while B	C after D	Is it possible that D caused A?	Cause after effect	0
A before B	D while B	C after D	Is it possible that A caused D?	Cause before effect	0
A before B	D while B	C after D	Is it possible that C caused A?	Cause after effect	1
A before B	D while B	C after D	Is it possible that A caused C?	Cause before effect	1

Appendix B. Problem schematics used in Experiment 2.

Premise 1	Premise 2	Premise 3	Premise 4	Question	Question structure	Consistency
A before B	C after B	D while C	A before C	Is it possible that D caused B?	Cause after effect	Consistent
A before B	C after B	D while C	A before C	Is it possible that B caused D?	Cause before effect	Consistent
A before B	C after B	D while A	C before A	Is it possible that C caused D?	Cause after effect	Inconsistent
A before B	C after B	D while A	C before A	Is it possible that B caused C?	Cause before effect	Inconsistent