Causation, the Law, and Mental Models

PHILIP N. **JOHNSON-LAIRD**
Princeton University

MONICA **BUCCIARELLI**
Turin University

ABSTRACT
The theory of mental models distinguishes between the meanings of causes and enabling conditions—a distinction not yet made in English and Continental law. The legal "but for" test (or sine qua non) for factual causes errs in both omission and commission. The theory of mental models therefore replaces it with more precise tests for "causes" and "enables". These and the rejection of "proximate" causes (also known as "legal" causes) would bring jurisprudence closer to the everyday concept of causation.

KEYWORDS
but-for test, causes, enabling conditions, mental models, proximate causes

CITATION
Johnson-Laird P.N, Bucciarelli M. 2023. *Causation, the Law, and Mental Models*, in Brigaglia M., Roversi C. (eds.), *Legal Reasoning and Cognitive Science: Topics and Perspectives*, «Diritto & Questioni pubbliche», Special Publication, August 2023.

DIRITTO & QUESTIONI PUBBLICHE / RECOGNISE
Legal Reasoning and Cognitive Science: Topics and Perspectives
Special Publication / August, 2023, pp. 81-102

ISSN 1825-0173

Causation, the Law, and Mental Models

PHILIP N. JOHNSON-LAIRD, MONICA BUCCIARELLI

> … the clarification of the structure of ordinary causal statements was and is an indispensable first step towards understanding the use of causal notions in the law
> HART & HONORÉ 1985, p. xxxiv

1. *Introduction*

A long-standing problem for the law is that its notions of causation differ from those of daily life. As jurists admit, these notions are also confusing. Here are three illustrative puzzles. An attacker stabbed an 18-year-old girl and she was taken to hospital. She needed an operation but she was a Jehovah's Witness, and refused a preliminary blood transfusion. She was told that she would die as a result, but she said she did not care. She died the next day. The court accepted that her refusal of the transfusion was the reason she died. Yet, they found her attacker guilty of manslaughter on the grounds that his stabbing caused her death[1]. Why isn't her not having the transfusion the cause of her death?

A man assaulted and raped a woman. She then took poison. He refused to summon medical help and imprisoned her for hours in a hotel room. She died a month later, and the court determined the cause was either the poison or its effects coupled with those of her wounds. It found the defendant guilty of murder. On appeal, the court maintained this verdict[2]. Granted that her death depended on the poison, why didn't her own action cause her death?

A plaintiff fell down an open elevator shaft in the defendant's building and was injured. The court decided that the "proximate" cause of the accident was that the door to the elevator was open, but that the defendant was not guilty of negligence, because no evidence showed that any employee of his had opened it, the plaintiff knew that the elevator sometimes crept up or down, and the open door was a clear danger[3]. How can an open door cause a person to fall?

We return to these cases at the end of this essay to show how it solves their puzzles. Law in the English-speaking world distinguishes between a *factual* cause and, as the previous example illustrates, a *proximate* (or *legal*) cause. Factual causes are supposed to be nothing more than everyday causes, but as we show proximate causes are mystifying. What is most extraordinary is that if you ask for the meaning of an assertion, such as:

The defendant's action caused the outcome

jurists seem unsure. Cognitive scientists sometimes avoid answering this question too, and

[1] *Regina v. Blaue* 61 Cr App R 271 (1975).
[2] *Stephenson v. State*, 179 N.E. 633 (Indiana. 1932).
[3] *Hope v. Longley*, 27 R.I. 579, 65 A. 300 (R.I. 1906).

when they do respond, the disparity in their views is vast. Their theories depend on modal logics of which there is a countable infinity, 'possible worlds', counterfactual conditionals, Bayesian probabilities, mechanisms, forces, powers, principles, and so on and on. The foundational mystery for this essay is therefore: what does the causal assertion above mean? Its meaning, we assume, determines what verifies the assertion, and whether an inference follows of necessity from the assertion. Hence, for example, part of the meaning of *cause* is not that it is an *unusual* event, because this inference does not follow of necessity:

The defendant's action caused the outcome.
Therefore, *the defendant's action was unusual.*

A further mystery is why the answer to our foundational question is not obvious. Our chapter presents solutions to both mysteries. It begins with a review of causation in the law, which demonstrates its self-confessed confusions. It then goes back to Aristotle and examines philosophical analyses of the meaning of causation. None will do. It introduces the theory of mental models—for which it marshals evidence—and it shows how this theory illuminates the concept of causation, leads to proper tests of causal relations, answers our fundamental question, and explains its mysterious difficulty. It concludes with five recommendations for legal practice.

2. *Legal concepts of causation*

Perhaps influenced by John Stuart MILL (1874) whom we'll come to by and by, common law in England and America, and the law in continental Europe, offer a "but for" test (the *sine qua non* test) to answer questions about factual causes, as in:

Is it true that but for *the defendant's action, the victim would not have died?*
If, and only if, it is true, then the defendant's action caused her death.

The test calls for an evaluation of what philosophers nowadays refer to as a "counterfactual" conditional: "If the defendant had not acted as he did, then the victim would not have died". Skeptics argue that it is difficult or impossible to verify counterfactual assertions. But, it is often feasible, as is evident in the many sports with counterfactual rules, e.g., golf, rugby, basketball, and others. In cricket, for instance, a batsman is out if the ball would have hit the wicket had it not hit his body without first hitting his bat. The umpire imagines the path of the ball had the player's body not blocked it, and makes the decision. An automatic TV system, Hawk-Eye, also computes the ball's counterfactual trajectory had it not hit the player's body.

Continental lawyers have proposed other analyses. For instance, Birkmeyer, a nineteenth century jurist, suggested that a cause was whichever condition had the greatest energy and so contributed most to the effect (HART & HONORÉ 1985, Ch. XV). But, continental law also often treats any condition necessary for an event as a cause, just as the *but-for* test does (HART & HONORÉ 1985, Ch. XVI). When we turn from legal principles to actual practice, courts allow for "intervening causes", for assignment of mutual negligence if it is impossible to determine which of several tortfeasors in fact caused damage[4], and sometimes that an action was the probable cause of an effect[5].

[4] *Summers v. Tice* (1948) 33 C2d 80. We thank Stuart Lichten for telling us about this case.
[5] *New York Times co. v. Suillivan* (1964) 376 U.S. 254.

In law, a person who causes harm is not necessarily liable for damages or for punishment. On this point, proximate (aka "legal") cause enters deliberations. It goes back at least to Sir Francis Bacon, the seventeenth century philosopher, and former Lord High Chancellor of England. He assumed that causation peters out over time (F. BACON 2013 [1630]) and so what matters is its immediate harm. MILL (1874, 240) echoes him. But, as the great expert on torts, Prosser, remarked: a proximate cause can be a matter of legal policy and have nothing to do with causation (PROSSER 1964). About proximate cause, he declared: «There is perhaps nothing in the entire field of law which has called forth more disagreement, or upon which the opinions are in such a welter of confusion» (PROSSER 1964, 311), a remark that is primary evidence for our judgment of conceptual turmoil about causation in the law. Yet, we do need to understand the policies and purposes proximate causes are supposed to serve. It aims to curtail the consequences of a person's actions that otherwise might be unforeseeable or continue forever. Perpetrators may also include those who should be protected from accountability. In such cases, the law eliminates liability, and it does so by determining that defendants—even those who caused injury—are not proximate causes, and not liable. At the very least, the concept is misnamed, and sometimes it distorts the concept of causation in order to eliminate liability. It demands spatio-temporal proximity between a cause and its effect. And so only the last in a series of human actors who do harm is its proximate cause (PROSSER 1964, 316). Granted that Brutus was the final conspirator to stab Caesar in Shakespeare's (or Francis Bacon's?) play, he is the proximate cause of Caesar's death. Of course, no causal justification exists to warrant this decision—Casca, who struck first, could have delivered the only mortal blow.

In summary, factual causes in law have an operational definition in the *but-for* test, yet as we will show it yields erroneous diagnoses. Proximate causes do not always concern "cause": their aim is to curtail liability, and so they can also do violence to causation. Likewise, no ground exists for jurists' anxiety that the ramifications of a causal action are boundless, so that proximate causes are needed to curtail them. When you turn off a tap, the water stops flowing, and that can be the end of a causal sequence. Our contention is that the problems of causation in the law have a two-fold solution. Proper policies that have nothing to do with causation should deal with exclusions of liability, and the *but-for* test should be replaced with one based on the correct analysis of factual causation. The first author put forward a skeletal version of this solution over twenty years ago (JOHNSON-LAIRD 1999). We hope that the consequences of the present account will outlast those of turning off a tap.

3. *Theories of causation*

A search for the right analysis of causation starts with Aristotle. He presents a brief typology of four sorts (see his *Metaphysics,* Book V, 1013a and b in ARISTOTLE 1984). Causes, he says, are origins. So, the substance of which something is made is its material cause, e.g., silver. The form in which it is made is its formal cause, e.g., the shape of a saucer. The origin of a change is its efficient cause, e.g., the silversmith who made the saucer. And the end for which something is done is its final cause, e.g., Athenian rituals calling for saucers. Aristotle allows that two things can be causes of one another, though not in the same way, e.g., health causes walking, and walking causes health (*ibidem*, 1013b). As often with Aristotle, the acuteness of his distinctions is striking. Yet, he does not attempt to answer our foundational question about the meaning of a causal assertion.

The first great thinker to try to answer the question was the philosopher David Hume, a pillar of the Scottish enlightenment after the rise of science. He wrote: «We may define a cause to be an object followed by another, and where all the objects, similar to the first, are followed by objects similar to the second. Or in other words, where, if the first object had not been, the second never had existed» (HUME 1988 [1748], 115). According to his first definition, a general causal assertion, such as:

Stabbing a person causes the person to bleed

has the following construal suitable for analysis in standard logic (see JEFFREY 1981):

If any person is stabbed then the person bleeds.

It refers to a constant conjunction of cause and effect. But, Hume's second definition, which he prefaces with "in other words," is different. It proposes that a singular event such as:

The stabbing of this man caused him to bleed

has a counterfactual interpretation:

If this man had not been stabbed then he would not have bled.

This analysis may be the origin of the *but-for* test (cf. MILL 1874, 237), i.e., *but for* his stabbing this man would not have bled.

After Hume, a deluge. Analyses of causality introduced temporal sequence, spatial contiguity, necessary and sufficient conditions, and scientific laws. As we have seen, some of these ideas entered the law. Yet, Hume anticipated and dismissed many of them. In daily life, however, one constraint is that the onset of a cause does not occur after the onset of its effect. The only violations are in accounts of time travel, and in a physicist's conjecture that there is only one electron that can travel backwards in time. Yet, cause and effect sometimes seem contemporaneous, as when a lead ball causes a dent in a cushion (KANT 1934 [1781], 156).

Spatial contiguity is not strictly necessary for causal relations in everyday life, as in psychological relations, e.g.:

Marcel's jealousy caused him to hate Albertine

or in Einstein's claim about "spooky" action at a distance in quantum physics (GOLDVARG & JOHNSON-LAIRD 2001). Spatio-temporal contiguity, however, can be a cue to a causal interpretation of events.

Other mooted transmitters of causation include a mechanism, power, force, energy, or means of production. KOSLOWSKI (1996, 6) wrote, «a causal mechanism is the process by which a cause brings about an effect». Clocks, engines, and computers, embody hierarchies of causal relations in hardware. Each relation may have its own underlying mechanism, but the "recursion" has to bottom out. So, at least one causal assertion does not invoke mechanisms, or else they exist "all the way down" like the turtles that were once said to support the earth. Our skepticism in no way impugns the role of known mechanisms in the induction of causal relations (e.g., AHN et al. 1995). Likewise, causes can transmit energy that has an effect (HARRÉ & MADDEN 1975, 5). A recent formulation roots causation in people's perceptions of force (WOLFF & BARBEY 2015). When a tennis player smashes a racket on the surface of the court, the energy of the blow is transferred to the racket, bending it out of shape. Other claims are more abstract, and call for scientific principles to lie behind every causal assertion about particular events (HART & HONORÉ 1985, Ch. I., Sec. II). The best evidence for all these accounts is that knowledge of a mechanism, power, force, or principle, can override the covariation of a putative cause and effect on judgments of causality (e.g., KOSLOWSKI 1996; WHITE 1995). Their presence is a cue to infer a causal relation, but, as we will show, their absence is not enough to infer its absence.

Theorists err when they confuse a philosophical assumption with an element of meaning. MILL (1874, 236) proposed a law of universal causation: «The truth that every fact which has a beginning has a cause». But, this law cannot be part of the meaning of "cause", because if it were, then those who dissent and argue, say, that the big bang had no cause, are not propounding an alternative philosophy, but arguing about the meaning of a word. Lexicography is not metaphysics. The assertion:

Every event has a cause

makes just as good sense as its denial, which is not the self-contradiction that Mill's law implies. The meaning of "causes" embodies neither his universal law nor its quantum mechanical denial. We draw an analogous moral about the crucial case of causation by omission, e.g.:

The signalman's failure to set the signal caused the crash.

MILL (1874, 239) recognized such causes, but argued: «From nothing, from a mere negation, no consequences can proceed». Hence, he claimed that omissions are causes only as a result of their positive effects. Proponents of the theory that causes depend on forces have made an analogous argument (WOLFF et al. 2010). They treat the assertion:

The lack of a jack caused the car to fall to the ground

as an instance of "double prevention". Hitherto, the force of the jack prevented the car from falling, but then another force—say, a mechanic kicking the jack away—prevented this prevention, and so the car fell to the ground. In general terms: A prevented B from preventing C, and so the kick (A) prevented the jack (B) from preventing the fall (C), and so the lack of B caused C. This account works for some assertions, but not for others, such as:

That there's no highest prime number caused Euclid an enormous surprise.

There are no forces or double preventions in this case. The force theory yields an insightful explanation of how interactions among forces yield causal relations. But, as Khemlani and his colleagues have shown, the theory of mental models—to be described below—provides a unified account of omissions as causes, as enabling conditions, and as preventions (KHEMLANI et al. 2018).

In sum, Occam's razor cuts the meanings of causal relations to their rudiments. It excises contiguity, power, force, energy, mechanism, means of production, and scientific principles. Knowledge can add these factors, but causal meanings themselves do not need them. Hume was right.

3.1. *Necessity and counterfactuals*

Is there a necessary connection between cause and effect, i.e., given a cause is its effect bound to occur? Hume dismissed necessity in place of constant connection. But, Kant argued that causation is a notion demanding «that something, A, should be of such a nature, that something else, B, should follow from it necessarily» (KANT 1934 [1781], 90). So, his claim depends on conceptual (or "alethic") necessity. Indeed, no difference exists in the conditions in which the following two assertions are true, or in those in which they are false, i.e., their "truth conditions":

These circumstances cause plants to grow.
Given these circumstances, plants are bound to grow.

Here, "bound to" plays a role akin to certainty or to necessity (JOHNSON-LAIRD & RAGNI 2019).
The second part of Hume's analysis of causation, which we cited earlier, relates it to counterfactual claims, such as:

If the cause hadn't happened then the effect wouldn't have happened.

Various modern philosophers have also advocated a counterfactual semantics for causation (e.g., LEWIS 1973a; MACKIE 1980). So, a particular causal claim, such as:

His lack of vitamin C caused him to develop scurvy

is synonymous with:

If he hadn't lacked vitamin C then he wouldn't have developed scurvy.

Such counterfactual conditionals have well-known semantic accounts based on "possible worlds" (e.g., LEWIS 1973b; STALNAKER 1968). Their essence is that the preceding conditional is true if the claim that he did not develop scurvy is true in the possible world or worlds most like the real world amongst those in which he did not lack vitamin C. Hence, his deficiency in vitamin C leads of necessity to scurvy. The number of possible worlds is too vast to be psychologically plausible, and so the Hawk-Eye TV system is a better model of how humans assess counterfactuals. They envisage how one situation would be without a key component, the lack of vitamin C, and they evaluate the alleged consequence in this modified situation. Yet, the moral remains the same: the truth of the counterfactual above implies the alethic necessity of the effect given the cause.
Counterfactuals are instantiated in PEARL's (2009) use of Bayes nets to explain causation (see also SPIRTES et al. 2000; SLOMAN et al. 2009). Pearl argues that causation is a deterministic notion. A purely probabilistic account, he says, cannot distinguish between *rain causes mud* and *mud causes rain*. Human ignorance, however, demands that causation is treated as probabilistic. The structure of "causal diagrams" is at the core of Pearl's approach. In one of his examples (PEARL 2009, Figure 7.1), a court orders the execution of a prisoner. The captain in charge of the firing squad gives the order to two riflemen. Both of them shoot, and the prisoner dies. Figure 1 presents this scenario in a causal diagram in which each node refers to a proposition that can either be true or false. Pearl's argument begins with an analysis based on standard logic in which each arrow in the diagram denotes a biconditional of the sort, *If and only if A then B*, which we abbreviate as, *Iff A then B*, and so the figure is equivalent to this set of assertions:

Iff the Court orders the execution then the captain gives the order to the riflemen.
Iff the captain gives the order to the riflemen then rifleman 1 fires.
Iff the captain gives the order to the riflemen then rifleman 2 fires.
Iff rifleman 1 fires or rifleman 2 fires, or both fire, then the prisoner dies.

In standard logic, a biconditional, *iff A then B*, is true when *A* & *B* is true or when *not A* & *not B* is true; otherwise, it is false (JEFFREY 1981, 71). It follows that Figure 1 represents only two contingencies: either all its assertions are true, or else all of them are false. So, various conclusions can be drawn from logic alone, such as:

If rifleman 1 didn't fire then the prisoner didn't die.

But, consider the counterfactual event in which rifleman 1 fires without waiting for the captain's order. It violates the logical interpretation of Figure 1. But, it is the sort of event that can occur. PEARL (2013) treats it as demanding an *intervention* that calls for surgery on the diagram in which a "do-operator" cuts the link from the captain's order to rifleman 1 in Figure 1. All else remains the same, and so the following counterfactual conditional is now true:

If rifleman 1 hadn't fired then the prisoner wouldn't have died.

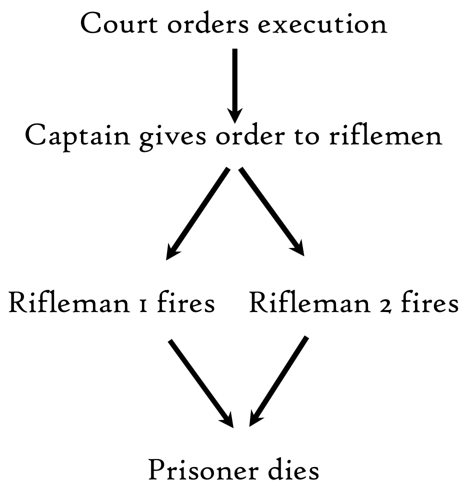

FIGURE 1. A causal diagram of the firing squad example (after PEARL 2009, Figure 7.1)

This treatment amounts to altering a representation of reality, like Hawk-Eye does, by removing one of its components.

Pearl assumes that causation depends on an intervention, not a mere observation. So, whenever an intervention occurs, you remove all arrows from the causal diagram that point to the event (rifleman 1 fires), you fix its value to true, and you use logic to deduce the consequences (the prisoner dies). Pearl's central principle is thus:

Y causes Z if we can change Z by manipulating Y (PEARL 2009, Ch. 11).

He does not intend this principle to define the meaning of "causes". Indeed, it doesn't, because it is not self-contradictory to assert:

The big bang caused the universe to begin, and we cannot manipulate the big bang.

Causal diagrams do not represent conditions that *enable* events to occur, but Sloman and his colleagues have shown how they might do so. The crucial part of *A enables B to happen* is «that A represents a category of events necessary for B, and that an alternative cause of B exists» (SLOMAN et al. 2009, 21). Yet, the lack of a cause does not contradict an enabling assertion, e.g., "The gas enabled an explosion to occur, but luckily no spark occurred to cause it". An enabling condition requires only the possibility of a cause, not its occurrence. As Pearl argues, his analysis supports the *but-for* test in law:

But for the rifleman firing, the prisoner wouldn't have died.

Every arrow in a causal diagram is supposed to signify a potential cause, and the theory presupposes that causation has no simpler analysis. By design, however, it offers no answer to our foundational question about the meaning of "causes".

The example of the firing squad elides an important distinction. The captain ordered the riflemen to fire, but orders are not causes. You cannot disobey a cause. The law reflects the difference: human beings can decide whether or not to obey to an order (HART & HONORÉ 1985, Ch. II). Like causes, words can embody intentions in their meanings, e.g.: chase, emigrate, examine, aim, look for, seek, assassinate (see MILLER & JOHNSON-LAIRD 1976, Sect. 6.3). You can't chase something by accident. Whether an action was intentional can affect a defendant's culpability, but not whether it is a cause.

3.2. *Probabilistic theories*

Uncertainty plagues causality. A philosophical view, originating in the twentieth century, is that the causation is itself probabilistic. REICHENBACH (1956) proposed such an analysis, arguing that *C causes E* if the probability of E given C is greater than the probability of E given not C. He allowed, however, that a putative cause could be "screened off" if there was an earlier probabilistic cause, D, yielding both C and E. Cognitive scientists have also defended probabilistic theories of the meaning of conditionals (e.g., OAKSFORD & CHATER 2007). And Cheng and her colleagues have defended a similar theory of causation based on normalizing the difference between the two conditional probabilities above (CHENG 1997). «Because causal relations are neither observable nor deducible, they must be induced from observable events» (CHENG 1997, 367). In fact, when a display appears to show one billiard ball colliding with another stationary one, it elicits a compelling perception of the first ball causing the second one to move, even though it is, in fact, an illusion (MICHOTTE 1963 [1946]).

The main case for a probabilistic semantics is that people judge that a causal relation holds in instances in which exceptions occur (e.g., CHENG & NOVICK 1991). But, this evidence is hardly decisive. People accept that the coronavirus causes Covid-19, yet they know that not everyone who has the virus develops the disease. Hidden variables affect a person's susceptibility to it. The claim itself is a generic one, akin to *ducks lay eggs*, which people accept, even though they know that only females lay eggs (LESLIE et al. 2011). It is therefore sensible to maintain that an assertion such as:

Smoking caused his death.

has a deterministic meaning, because it differs from:

Smoking probably caused his death.

If causation were probabilistic, the two assertions should share the same truth conditions. Pearl is right: causation is not probabilistic, but the evidence supporting it can be. So, causation and probabilistic considerations need to be kept separate—a division that many theories appear to recognize (see, e.g., WALDMANN 1996; LAGNADO & GERSTENBERG 2017).

3.3. *Causes and conditions*

A spark occurs in a container of combustible gas, and an explosion follows. *But for* the co-occurrence of gas and spark, there would have been no explosion. Hence, MILL (1874, 238) treated both as "conditions", and argued that the choice of one of them to be the cause was capricious. Other theorists have argued that causes are abnormal whereas enabling conditions are normal (HART & HONORÉ 1985, Ch. II), that causes are inconstant whereas enabling conditions are constant (CHENG & NOVICK 1991), that causes violate a norm whereas enabling conditions occur by default (e.g., EINHORN & HOGARTH 1986), and that causes are relevant to explanations whereas enabling conditions are not (MILL 1874, 238; MACKIE 1980; TURNBULL, SLUGOSKI 1988; HILTON & ERB 1996). What all these accounts presume is that there is no difference in meaning between causes and enabling conditions. Readers might suppose that, say, *abnormal* is part of the meaning of cause. But, as assertion such as: "A spark is not an abnormal cause of an explosion", makes perfect sense, whereas an assertion such as: "A spark occurring after the explosion caused it" is nonsensical. Such tests show that inconstancy, violation of a norm, and relevance to explanation, are also not part of the meaning of "cause". The theory of mental models, to which we now turn, makes the same point.

4. *The mental model theory*

4.1. *The basics*

Mental models are representations of the world, and the development of the "model" theory—as it is referred to—has accumulated five main principles:

(1) Each mental model represents a possibility with many different instances that have in common only what the model represents. Its structure is *iconic* in that it matches the structure of what it represents insofar as possible. So, a kinematic model unfolds in time, just as the sequence of events that it represents does (JOHNSON-LAIRD 1983; KHEMLANI et al. 2013). Models can also contain abstract symbols, which are not iconic, such as negation, and each of them is linked to its semantics.

(2) Individuals cope best with one model at a time (JOHNSON-LAIRD 1983). An *intuitive* model represents only those clauses in the premises that are true; but a *deliberative* model also represents those clauses that are false—using negation to do so. So, humans have an intuitive system of reasoning, and a deliberative system—an idea due to the late Peter Wason, which has since flourished in many forms, though only the model theory appears to have implemented the two systems in a computer program. Intuition focuses on a single model at a time; deliberation can find models that serve as counterexamples to intuitive conclusions (e.g., KHEMLANI & JOHNSON-LAIRD 2021; RAGNI et al. 2018).

(3) Models represent three main sorts of possibility (JOHNSON-LAIRD & RAGNI 2019): what can happen (epistemic possibilities), what is permissible (deontic possibilities), and what is conceptually feasible (alethic possibilities). They can each refer to real or to counterfactual possibilities, i.e., situations that were once possible but that didn't happen (BYRNE 2005; BYRNE & JOHNSON-LAIRD 2020). The manipulation of a model of reality can assess the veracity of a counterfactual (GERSTENBERG et al. 2019; BYRNE et al. 2022).

(4) The interpretation of any assertion is open to *modulation* in which knowledge, meanings, and referents, can eliminate models, add information to them, or assign a truth value to them (e.g., QUELHAS et al. 2019).

(5) An inference is alethically necessary if its conclusion holds only in the models of its premises. Reasoners withdraw a conclusion when they discover that it is false, and revise the premises to try to construct an explanation that resolves the inconsistency (JOHNSON-LAIRD et al. 2004).

Mental models differ from standard logic in several ways. In logic, a contradiction yields valid inferences of any conclusion whatsoever, because there are no cases in which the premises are true and the conclusion is false (JEFFREY 1981, 1). In mental models, contradictory premises yield an empty model from which it follows only that something is wrong with the premises. Its effects are local. As HINTERECKER and colleagues (2016) showed, people accept inferences of this sort:

Either she died from poison or she died from her wounds.
So, *it is possible that she died from poison.*

In the model theory, the premise yields a conjunction of possibilities, which each holds in default of knowledge to the contrary. One of these possibilities is the conclusion above. Yet it is invalid in all standard modal logics dealing with possibilities.

The model theory applies to all sorts of reasoning, and it explains how reasoners' emotions can improve their reasoning (GANGEMI et al. 2013). It has been implemented in computer programs that take descriptions as their inputs, build models of the situations under description, and use these models both to draw conclusions—just as human reasoners do—and to evaluate given conclusions[6]. We now consider what the theory says about causation.

4.2. *The meanings of general causal and enabling assertions*

Models can represent forces, mechanisms, spatial and temporal contiguity, and much else. But, one of the theory's principal assumptions, going back to GOLDVARG & JOHNSON-LAIRD (2001) and from thence to MILLER & JOHNSON-LAIRD (1976, Sect. 6.3) is:

> «The meanings of causal relations refer solely to sets of possibilities with the temporal constraint that causes do not start after their effects do» (JOHNSON-LAIRD & KHEMLANI 2017, TABLE 10.1).

Any causal assertion is synonymous with a conditional assertion referring to epistemic possibilities (JOHNSON-LAIRD & BYRNE 2002; BYRNE & JOHNSON-LAIRD 2020). For instance, the causal assertion:

Taking substance D causes you to sleep

has the same truth conditions as the conditional assertion:

If you take substance D then you sleep.

It does not follow that all conditionals express causal relations. They can assert other sorts of relation, such as so-called "evidentials", e.g., *If the door is closed, then the elevator is on another floor*. And causal assertions themselves can be elliptical, e.g., *Drivers cause accidents*, implies that the actions or inactions of drivers cause accidents to occur.

The meaning of a general causal assertion refers to a conjunction of default possibilities that have the appropriate temporal constraint. Individuals tend to represent possibilities in *intuitive* models of the world, which the computer program implementing the theory denotes using words. So, the causal claim that taking substance D causes you to sleep has these two intuitive models:

substance D	sleep
. . .	

The first row in this diagram designates a model of the salient possibility in which you take

6 These programs are at https://www.modeltheory.org/models/.

substance D and sleep, and the second row, the ellipsis, designates a model with no explicit content but that allows for alternative possibilities in which you do not take substance D. When people think hard, they can enumerate these possibilities in *deliberative* models, where "not" is a symbol for negation:

	substance D		sleep
not	substance D	not	sleep
not	substance D		sleep

The first model represents the possibility in which a person takes substance D and sleeps; the second model represents the possibility in which the person does not take substance D and does not sleep; and the third model represents the possibility in which the person does not take substance D and sleeps. The three models establish that substance D causes you to sleep. It suffices, but, as the third model shows, it is not necessary, because other causes of sleep are possible too. What the assertion rules out as impossible is that you take substance D and do not sleep.

Some causes are unique, such as:

Drinking excessive alcohol causes you to get drunk.

Excessive alcohol is the only cause of becoming drunk: it suffices and it is necessary. The assertion refers to only two possibilities, which have these deliberative models:

	excessive alcohol		drunk
not	excessive alcohol	not	drunk

Enabling assertions refer to a different set of possibilities to those for causes. For instance, the assertion:

Taking substance D enables you to sleep

has these intuitive models:

substance D	sleep
. . .	

and so the possibilities seem the same as those for a causal assertion. However, the deliberative models of enabling conditions differ from those above for causes:

	substance D		sleep
	substance D	not	sleep
not	substance D	not	sleep

Substance D is necessary for you to sleep—without it, you won't sleep, but with it, you may or may not sleep. A weaker sort of enabling condition is one that also allows you to sleep without substance D: its deliberative models represent all four contingencies as possible.

To prevent something is to cause it not to occur, and prevention too can be unique. So, the models for:

Taking substance E prevents you from sleeping are identical to those for:

Taking substance E causes you not to sleep.

The model theory allows that a non-event can cause an effect (KHEMLANI et al. 2018). So, for example, the assertion:

Not taking substance E causes you to sleep

refers to these possibilities:

not	substance E		sleep
	substance E	not	sleep
	substance E		sleep

The second and third possibilities show that on taking substance E, you may or may not sleep. This example could be a double prevention, which we described earlier (WOLFF et al. 2010). But, other examples cannot be:

The idea that Zeus does not exist caused ancient Athenians to laugh.

It is nonsensical to suppose that Zeus's existence prevented Athenians from laughing.

TABLE 1 presents the four main sorts of causal assertion, and their intuitive and deliberative models. Each assertion refers to a conjunction of a set of possibilities that each hold in default of knowledge to the contrary. Causes and preventions can be unique, and therefore have only two models. Enabling conditions can be weak and therefore have all four models. The variables A and B denote events or situations, but they can also denote their non-occurrence or non-existence. The complete set of causal relations therefore consists of seven relations: the four sets of three possibilities in TABLE 1, two sets of two possibilities for strong causes and preventions, and one set of all four possibilities for a weak enabling condition (see JOHNSON-LAIRD & KHEMLANI 2017, for a proof that these exhaust the set of possible causal relations).

4.3. *Empirical corroborations*

Several lines of evidence have corroborated the model theory of causal relations. One line bore out the difference in meaning between causes and enabling conditions. When participants in an experiment had to list what is possible and what is impossible given assertions based on each of the main sorts of causal verb—*causes*, *enables*, and *prevents*, their listings had reliable matches with the predicted possibilities (GOLDVARG & JOHNSON-LAIRD 2001, Experiment 1). A further study showed that individuals distinguish between causes and enablers when they occur together in a ternary relation (GOLDVARG & JOHNSON-LAIRD 2001, Experiment 2). The participants were given brief scenarios, and their task was to identify the cause: "it brings about the event," and the enabling condition: "it makes the event possible". Here is a typical scenario:

Given that there is good sunlight, if a certain new fertilizer is used on poor flowers, then they grow remarkably well. However, if there is not good sunlight, poor flowers do not grow well even if the fertilizer is used on them.

Causal Assertions	Intuitive Models	Deliberative Models
A causes B.	A B . . .	A B not A not B not A B
A enables B.	A B . . .	A B A not B not A not B
A prevents B.	A not B . . .	A not B not A B not A not B
A enables not B.	A not B . . .	A not B A B not A B

TABLE 1. The four principal sorts of causal relation, the intuitive models of their possibilities, and the deliberative models of all their explicit possibilities. The ellipsis denotes an intuitive model with no explicit content, and the variables, A and B, can also have negative instances as their values. Strong interpretations of *causes* and *prevents* refer to only their first two possibilities in the table; weak interpretations of *enables* and *enables not* refer to all four possibilities.

As the participants realized, with sunlight, the fertilizer causes growth; without sunlight, it does not. The ternary relation encapsulating the scenario is:

Sunlight enables the fertilizer to cause the flowers to grow.

The experiment included scenarios embodying a ternary relation switching the roles of the two agents:

The fertilizer enables sunlight to cause the flowers to grow.

It also counterbalanced the order of mention of the causers and enablers, and all the participants saw different contents for each of the four different sorts of scenario. They correctly identified causes and enabling conditions on 85% of occasions. The switch in causal roles between fertilizer and sunlight refutes the role of abnormality, rarity, violation of norms, and other such concepts, as essential to causes. The pattern of possibilities for all eight cases for fertilizer, sunlight, and growth differs between the two assertions above (see GOLDVARG & JOHNSON-LAIRD 2001, 578 f.).

A similar study of the legal implications of causes and enablers used scenarios about harmful effects (FROSCH et al. 2007). A typical scenario was:

Mary threw a lighted cigarette into a bush. Just as the cigarette was going out, Laura deliberately threw petrol on it. The resulting fire burnt down her neighbor's house[7].

[7] Compare: *Watson V. Kentucky* & *Indiana Bridge* & *Railway Co.* (1910) 137 KY. 619, 126 S.W. for the doctrine of "intervening cause"—a concept that Stuart Lichten told us about. The initial act of negligence is curtailed by the intervening causal action, but only if it is intentional.

Once again, participants distinguished between causers and enablers. They judged causers as more responsible for the consequences than enablers, and as deserving to serve much longer prison sentences and to pay much greater damages. The causers' responsibility, as a recent study showed, is judged to be even greater when the order of events is described in the same order as they occur in reality than in the opposite order (LIMATA et al. 2022). The natural order of the enabling condition as occurring prior to the cause may make the construction of a kinematic model easier, and so the difference in responsibility between the two agents is clearer. As readers should recall, neither Anglo-American nor continental law recognizes the difference in meaning between a causer and an enabler. Yet, it affects culpability in daily life.

Another line of evidence for the model theory concerns the consequences of causes and enabling conditions. Most participants in a study responded "yes" to the following sort of inference (GOLDVARG & JOHNSON-LAIRD 2001, Experiment 1):

Eating protein will cause her to gain weight.
She will eat protein.

Will she gain weight?
But, as the models of their possibilities in TABLE 1 predict, they responded "no" to the same inference when its initial premise referred to an enabling condition:

Eating protein will enable her to gain weight.

The assertion allows that she may not gain weight.

Still another line of evidence bore out the model theory's deterministic meaning for causal relations (FROSCH & JOHNSON-LAIRD 2011). In various experiments, individuals enumerated the evidence that they thought would refute general causal claims. Most participants called for single pieces of evidence to refute causes, enablers, and preventions. They tended to choose refutations of the form, *A and not B*, for both causes and enablers, but, as predicted, they chose those of the form: *not-A and B* more often for *A enables B* than for *A causes B*.

In sum, contrary to the tradition that MILL (1874) inaugurated, causal and enabling conditions differ in meaning, in inferential consequences, and in evidence that refutes them. Individuals can tell them apart. And the difference does not depend on normal versus abnormal conditions, constant versus inconstant conditions, probable versus improbable conditions, or on what is relevant versus irrelevant to explanations. The contents in the preceding experiments rule out these factors as crucial distinctions. All that matters are the possibilities to which assertions refer and their appropriate temporal constraint.

4.4. *Novel causal relations*

The model theory distinguishes between the possibilities to which various ternary relations refer, such as: *A enables B to cause C*, and *A causes B to cause C*. Their possibilities differ from those for the conjunctions of relations: *A enables B, and B causes C*, and *A causes B, and B causes C* (GOLDVARG & JOHNSON-LAIRD 2001). The theory allows for two-way causal relations—a pertinent example is that people's beliefs about deontic matters affect their emotions, and their emotions about these propositions affect their beliefs in them (BUCCIARELLI & JOHNSON-LAIRD 2019a, 2019b, 2020; BUCCIARELLI et al. 2008). The theory also allows for cyclical causal relations, as in feedback loops in the body and in other systems. It is sensitive to the difference between causes and orders from persons in authority. No other theory, as far as we know, copes with all of these matters.

5. *The legal implications of the model theory*

The principal implications of the model theory for jurisprudence are to refute the *but-for* test, which makes errors of commission and omission, to establish viable replacements for it, and to elucidate hitherto puzzling legal judgments.

The essential problem with the *but-for* test is that it establishes that one event is necessary for the occurrence of another.

But for his failure to wear a seat-belt, he would not hit his head on the dashboard.

Granted that the assertion is true, his failure to wear a seat-belt did not cause him to his head on the dashboard. It enabled the event to happen. An enabling condition is necessary for an event to occur, because without it, the event does not occur (see the possibilities in TABLE 1). Hence, the test works for unique causes, because they are necessary (and sufficient) for the effect to occur. But, it makes an error of commission in treating enabling conditions as causes.

Consider this claim in a situation with a firing squad:

The rifleman's shot caused the prisoner's death.

The three prior possibilities are as follows (see the possibilities in TABLE 1):

rifleman shoots	prisoner dies
not rifleman shoots	not prisoner dies
not rifleman shoots	prisoner dies

The third possibility is that another rifleman shoots and the prisoner dies. The fact described in the assertion above converts the first of the three possibilities into a fact, and the second and third of them into counterfactual possibilities; what remains impossible, as in Pearl's scenario above, is that the rifleman shoots and the prisoner does not die. The assertion above therefore makes a true causal claim, but the but-for test fails to diagnose it:

But for the rifleman shooting, the prisoner would not have died.

The test assertion is false, because another rifleman could have shot the prisoner. So, the *but-for* test also makes errors of omission. It fails to detect certain causes that are sufficient but not necessary. It is remarkable that a great philosopher should have envisaged this flawed test, and that so many eminent scholars have endorsed it. The reason is that unaided thought cannot hold in mind at the same time four different contingencies (JOHNSON-LAIRD 1983). We therefore propose to replace the *but-for* test with two new tests, one for causes, and one for enabling conditions. They derive from the sets of possibilities in TABLE 1.

5.1. *The cause-test and the enable-test*

The *cause*-test determines whether *A caused B* given an observation of their conjunction in an appropriate temporal order. The test does double duty and diagnoses prevention given *A*, and *B* is a non-occurrence. Figure 2 presents the *cause*-test, and we illustrate how it works using an example:

Laura threw petrol on a smouldering cigarette and a fire occurred.

We use common sense and knowledge to test whether Laura's action (A) of throwing petrol on a smouldering cigarette caused the fire (B), and the test allows only a small change to the actual situation in the evaluation of counterfactuals. The answer to the first question in Figure 2 is that Laura's action (A) with no subsequent fire (not B) could not have occurred. The answer to the second question is: Yes, Laura not carrying out her action (not A), with no subsequent fire (not B) could have occurred. The answer to the third question is: No, without her action, no fire could have occurred. The resulting diagnosis is therefore:

Laura's action caused the fire to occur.

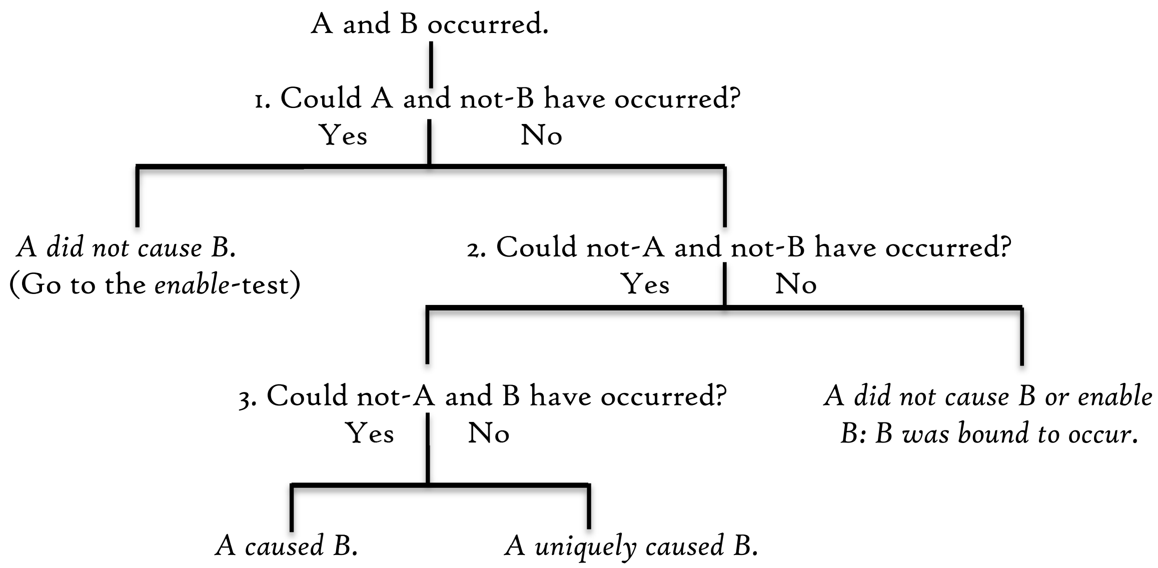

FIGURE 2. The *cause*-test: given the conjunction of A and B in an appropriate temporal order, the answers to three questions diagnose whether *A caused B*. The variables, A and B, each refer either to situations or to their non-occurrences, so when B is a non-occurrence, *causes* is equivalent to *prevents*.

Figure 3 presents the *enable*-test, which determines whether *A enabled B* given their appropriate conjunction. It too depends on the answers to three questions, and we illustrate it using a different example:

Mary threw a cigarette into a bush and a fire occurred.

The answer to the first question in Figure 3 is: Yes, Mary's action (A) could have occurred without the subsequent fire (not B): the smouldering cigarette could have gone out. The answer to the second question is: No, her not carrying out her action (not A) but the fire still occurring (B) couldn't have occurred (without a major change to the situation). The answer to the third question is: Yes, she could have not carried out her action (not A) and the fire not occurred (not B). The resulting diagnosis is:

Mary's action enabled the fire to occur.

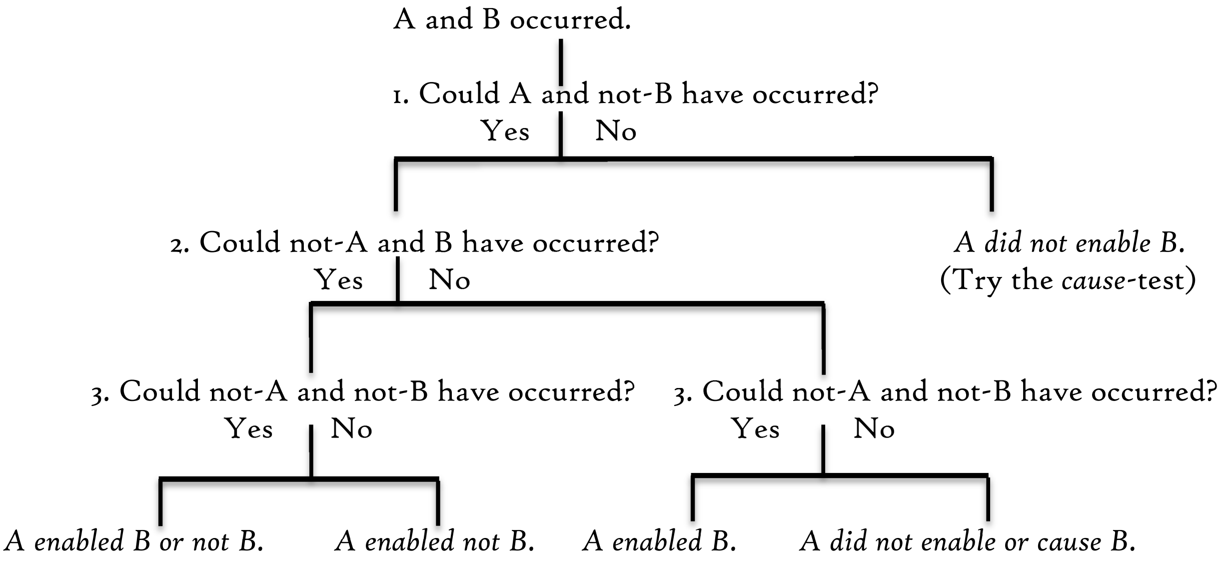

FIGURE 3. The *enable-test*: given the conjunction of A and B in an appropriate temporal order, the answers to three questions about diagnose whether A enabled B. The variables, A and B, each refer either to situations or to their non-occurrences, so when B is a non-occurrence, *enables* is equivalent to *enables B not to occur*.

The tests are more complex than the *but-for* test, but they deliver the right results.

5.2. *An elucidation of the courts' judgements in the three test cases*

The model theory's *cause*-test and *enable*-test elucidate the three court cases, which began this article. The man who stabbed the Jehovah's Witness caused her wounds. Their treatment called for a surgical operation. The court decided that her not having the preliminary blood transfusion for the operation was the "reason" for her death, but assigned its cause to her wounds. The answers to the *cause*-test (where A = No transfusion or operation, and B = death) are:

No transfusion	death:	Fact
No transfusion	no death:	Question 1: No: it could not have occurred.
transfusion	no death:	Question 2: Yes: it could have occurred.
transfusion	death:	Question 3: Yes: it could have occurred.

So, not having the blood transfusion and operation caused her death. Yet, the court judged her wounds as the cause of her death; they enabled it to occur. The court therefore erred.

The case of the woman who was attacked and raped was different. Her terrible assault was the reason she poisoned herself. As the court decided, the cause of her death was either the poison alone or else its combination with her wounds. The *cause*-test on the poison yields the answers:

poison	died:	Fact
poison	did not die:	Question 1: No: it could hot have happened.
no poison	did not die:	Question 2: Yes: it could have happened.
no poison	died:	Question 3: Yes: it could have happened.

So, her own action caused her death. In contrast, the *cause*-test on her wounds fails the first question: it could have happened that with her wounds she did not die. Hence, the poison with or without her wounds caused her death. Her assailant caused her death only if his attack or his delaying her treatment, or both, had mortal consequences. His actions were the reason for her intentional action of taking the poison.

In the case of the man who fell down an elevator shaft, the court decided that the proximate cause of his accident was the open door. The *cause* and *enable* tests show that the open door did not cause his fall, because he might not have fallen even though the door was open. Instead, it enabled him to fall: had it been closed, he could not have fallen. The court mistook an enabling condition for a cause. They could then judge that the defendant was not guilty of negligence. The *enable*-test yields these results:

Open door	injured:	Fact
Open door	not injured:	Question 1: Yes: it could have happened.
Closed door	injured:	Question 2: No: impossibility
Closed door	not injured:	Question 3: Yes: it could have happened.

So, the open door enabled the man to fall and to injure himself.

We have explored the application of the two tests to complex scenarios characterized by a ternary relationship in which one action enables another to function as a cause. So, when the enabler occurs, the second event functions as a cause; and when the enabler does not occur, the second event cannot function as a cause. This relation is distinct from the conjunction of A enables B, and B causes C (GOLDVARG, JOHNSON-LAIRD 2001), and a typical instance occurs when a catalyst enables a chemical to have a causal effect on another. Future experiments could explore the possibility of developing a test that allows this ternary relationship to be explicitly evaluated, and to be contrasted with the conjunction of the two relations. What makes such tests seem impractical is the difficulty of holding distinct models in mind at the same time.

6. *General Discussion*

The model theory explains the meanings of causal relations, their mental representations, and inferences from them. It postulates only rudimentary meanings for causal relations. With an appropriate temporal order: *A causes B* means that given *A* the occurrence of B is the only possibility, *A enables B* means that given *A* the occurrence of *B* is one possibility, and *A prevents B* means that given *A* the occurrence of not-*B* is the only possibility[8]. These relations are deterministic. If they were probabilistic then *causes* would tolerate exceptions and so it would be indistinguishable from *enables*. Actual legal cases contain many hidden factors, which may be alternative causes, enabling conditions, or preventions, and so diagnosis can be difficult. Alternative theories of causation base their interpretations on forces, powers, mechanisms, and interventions. The model theory accommodates these elements: models can embody them as a result of modulation from knowledge and context. Yet, non-events can be causes, which demonstrate that these factors are not part of the fundamental meanings of causal claims.

What can jurists take away from the model theory's account of causation? In our view, it yields five lessons:

[8] See TABLE 1.

(1) The fundamental assertion:

The defendant's action caused the outcome

has two potential meanings, referring either to one among a set of causes or to a unique cause: the defendant carried out an action with an onset no later than the outcome's onset, and given this action the outcome was bound to occur. Without this action, if the outcome might still have occurred, the cause is just one of several; otherwise, the cause is unique.

(2) Both *A causes B* and *A enables B* have the same intuitive models of a conjunction of default possibilities, one explicit and the other with no explicit content:

A B
. . .

The identity of their intuitive models explains past failures to distinguish between causes and enabling conditions. Only deliberative models, and the *cause*-test and *enable*-test, which make all the possibilities explicit, differ between the two[9].

(3) Contiguity in space and time, force, energy, power, and mechanism, are cues to causation, and modulation can embody them in models. But, they are not part of its essential meaning, because they cannot occur in omissions that are causes, enabling conditions, or preventions.

(4) Subtleties can affect legal judgments: an order from a person in authority can be reason for an action, but not its cause; relations can be ternary so that one event enables another to cause an effect; causal relations can be two way and form feedback loops.

(5) The legal *but-for* test is flawed, and errs both in omitting genuine causes and in treating enabling conditions as causes. The *cause*-test and *enable*-test are better guides to diagnosis in real life.

In conclusion, causation is a matter of possibilities, and possibilities can relate to one another in complex ways. If legal decisions distinguished between the culpability of enabling harm and causing it, replaced the *but-for* test with the *enable*-test and the *cause*-test, and abandoned proximate causes in favor of principles curtailing liability, they would take an indispensable first step towards causation as people conceive it in daily life.

[9] See TABLE 1.

References

ANH W., KALISH C.W., MEDIN D.L., GELMAN S.A 1995. *The Role of Covariation Versus Mechanism Information in Causal Attribution*, in «Cognition», 54, 299 ff.

ARISTOTLE 1984. *Metaphysics*, in ID., *The Complete Works of Aristotle*, Vol. 1 (ed. by J. Barnes), Princeton University Press.

BACON F. 2013. *Maxims of the Law*, Cambridge University Press. (Originally published 1630).

BUCCIARELLI M., JOHNSON-LAIRD P.N. 2019a. *Deontics: Meaning, Reasoning, and Emotion*, in «Materiali per una storia della cultura giuridica», 49, 89 ff.

BUCCIARELLI M., JOHNSON-LAIRD P.N. 2019b. *Emotions and Beliefs about Morality can Change one Another*, in «Acta Psychologica», 198, 102880.

BUCCIARELLI M., JOHNSON-LAIRD P.N. 2020. *Beliefs and Emotions about Social Conventions*, in «Acta Psychologica», 110, 103184.

BUCCIARELLI M., KHEMLANI S., JOHNSON-LAIRD P.N. 2008. *The Psychology of Moral Reasoning*, in «Judgment and Decision Making», 3, 121 ff.

BYRNE R.M.J. 2005. *The Rational Imagination: How People Create Alternatives to Reality*, MIT Press.

BYRNE R.M.J, JOHNSON-LAIRD P.N. 2020. *If and Or: Real and Counterfactual Possibilities in their Truth and Probability*, in «Journal of Experimental Psychology: Learning, Memory, and Cognition», 46, 760 ff.

BYRNE R.M.J., KHEMLANI S., JOHNSON-LAIRD P.N. 2022. *Mental Simulations of Counterfactual Possibilities Qualify Judgments of Truth and Falsity*, in submission.

CHENG P.W. 1997. *From Covariation to Causation: A Causal Power Theory*, in «Psychological Review», 104, 367 ff.

CHENG P.W., NOVICK L.R. 1991. *Causes versus Enabling Conditions*, in «Cognition», 40, 83 ff.

EINHORN H.J., HOGARTH R.M. 1986. *Judging Probable Cause*, in «Psychological Bulletin», 99, 3 ff.

FROSCH C.A., JOHNSON-LAIRD P.N. 2011. *Is Everyday Causation Deterministic or Probabilistic?* In «Acta Psychologica», 137, 280 ff.

FROSCH C.A., JOHNSON-LAIRD P.N., COWLEY M. 2007. *It's not my Fault, your Honor, I'm only the Enabler*, in «Proceedings of the 29th Annual Meeting of the Cognitive Science Society», 1755.

GANGEMI A., MANCINI F., JOHNSON-LAIRD P.N. 2013. *Models and Cognitive Change in Psychopathology*, in «Journal of Cognitive Psychology», 25, 157 ff.

GERSTENBERG T., GOODMAN N.D., LAGNADO D.A., TENENBAUM J.B. 2021. *A Counterfactual Simulation Model of Causal Judgments for Physical Events*, in «Psychological Review», 128, 936.

GOLDVARG Y., JOHNSON-LAIRD P.N. 2001. *Naive Causality: A Mental Model Theory of Causal Meaning and Reasoning*, in «Cognitive Science», 25, 565 ff.

HARRÉ R., MADDEN E.H. 1975. *Causal Powers*, Blackwell.

HART H.L.A., HONORÉ T. 1985. *Causation in the Law* (2nd Ed.), Oxford University Press.

HILTON D.J. ERB H.P. 1996. *Mental Models and Causal Explanation: Judgements of Probable Cause and Explanatory Relevance*, in «Thinking & Reasoning», 2, 273 ff.

HINTERECKER T., KNAUFF M., JOHNSON-LAIRD P.N. 2016. *Modality, Probability, and Mental Models*, in «Journal of Experimental Psychology: Learning, Memory, and Cognition», 42, 1606 ff.

HUME D. 1988. *An Enquiry Concerning Human Understanding*, in FLEW A. (ed.), Open Court. (Originally published 1748).

JEFFREY R. 1981. *Formal Logic: Its Scope and Limits* (2nd Ed.), McGraw-Hill.

JOHNSON-LAIRD P.N. 1983. *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness,* Harvard University Press.

JOHNSON-LAIRD P.N. 1999. *Causation, Mental Models, and the Law*, in «Brooklyn Law Review», 65, 67 ff.

JOHNSON-LAIRD P.N., BYRNE R.M.J. 2002. *Conditionals: A Theory of Meaning, Pragmatics, and Inference*, «Psychological Review», 109, 646 ff.

JOHNSON-LAIRD P.N., GIROTTO V., LEGRENZI P. 2004. *Reasoning from Inconsistency to Consistency*, in «Psychological Review», 111, 640 ff.

JOHNSON-LAIRD P.N., KHEMLANI S.S. 2017. *Mental Models and Causation*, in WALDMANN M.R. (ed.) *The Oxford Handbook of Causal Reasoning*, Oxford University Press, 169 ff.

JOHNSON-LAIRD P.N., RAGNI M. 2019. *Possibilities as the Foundation of Reasoning*, in «Cognition», 193, 130950.

KANT I. 1934. *Critique of Pure Reason* (trans. by J.M.D. Meiklejohn), Dutton. (Originally published 1781).

KHEMLANI S.S., JOHNSON-LAIRD P.N. 2021. *Reasoning about Properties: A Computational Theory*, «Psychological Review», available at: https://doi.org/10.1037/rev0000240.

KHEMLANI S.S., MACKIEWICZ R., BUCCIARELLI M., JOHNSON-LAIRD P.N. 2013. *Kinematic Mental Simulations in Abduction and Deduction*, in «Proceedings of the National Academy of Sciences», 110, 16766 ff.

KHEMLANI S.S., WASYLYSHYN C., BRIGGS G., BELLO P. 2018. *Mental Models and Omissive Causation*, in «Memory & Cognition», 46, 1344 ff.

KOSLOWSKI B. 1996. *Theory and Evidence: The Development of Scientific Reasoning*, MIT Press.

LAGNADO D.A., GERSTENBERG T. 2017. *Causation in Legal and Moral Reasoning*, in WALDMANN M.R. (ed.), *The Oxford Handbook of Causal Reasoning*, Oxford University Press, 565 ff.

LESLIE S.J., KHEMLANI S.S., GLUCKSBERG S. 2011. *Do all Ducks Lay Eggs? The Generic Overgeneralization Effect*, in «Journal of Memory and Language», 65, 15 ff.

LEWIS D. 1973a. *Causation*, in «Journal of Philosophy», 70, 556 ff.

LEWIS D. 1973b. *Counterfactuals*, Harvard University Press.

LIMATA T., IANÌ F., BUCCIARELLI M. 2022. *Story Order in Attribution of Moral Responsibility*, in «Language and Cognition», 14, 228 ff.

MACKIE J.L. 1980. *The Cement of the Universe: A Study in Causation* (2nd Ed.), Oxford University Press.

MICHOTTE A. 1963. *The Perception of Causality*, Methuen. (Originally published 1946)

MILL J.S. 1874. *A System of Logic, Ratiocinative and Inductive: Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation* (8th Ed.), Harper.

MILLER G.A. JOHNSON-LAIRD P.N. 1976. *Language and Perception*. Cambridge, MA: Belknap, Harvard University Press.

OAKSFORD M. CHATER N. 2007, *Bayesian Rationality*, Oxford University Press.

PEARL P. 2009. *Causality: Modals, Reasoning, and Inference* (2nd Ed.), Cambridge University Press.

PEARL J. 2013. *Structural Counterfactuals: A Brief Introduction*, in «Cognitive Science», 37, 977 ff.

PROSSER W.L. 1964. *Handbook of the Law of Torts* (3rd Ed.), St. Paul West.

QUELHAS A.C., RASGA C., JOHNSON-LAIRD P.N. 2019. *The Analytic Truth and Falsity of Disjunctions*, in «Cognitive Science», 43, e12739.

RAGNI M., KOLA I. JOHNSON-LAIRD P.N. 2018. *On Selecting Evidence to Test Hypotheses*, in «Psychological Bulletin», 144, 779 ff.

REICHENBACH H. 1956. *The Direction of Time*, University of California Press.

SLOMAN S., BARBEY A.K. HOTALING J.M. 2009. *A Causal Model Theory of the Meaning of Cause, Enable, and Prevent*, in «Cognitive Science», 33, 21 ff.

SPIRITES P., GLYMOUR C., SCHEINES R. 2000. *Causation, Prediction, and Search* (2nd Ed.), MIT Press.

STALNAKER R.C. 1968. *A Theory of Conditionals*, in RESCHER N. (ed.), *Studies in Logical Theory*, Basil Blackwell, 98 ff.

TURNBULL W., SLUGOSKI B.R. 1988. *Conversational and Linguistic Processes in Causal Attribution*, in Hilton D. (ed.), *Contemporary Science and Natural Explanation: Commonsense Conceptions of Causality,* Harvester Press, 66 ff.

WALDMANN M.R. 1996. *Knowledge-Based Causal Induction*, in «Psychology of Learning and Motivation», 34, 47 ff.

WHITE P.A. 1995. *Use of Prior Beliefs in the Assignment of Causal Roles: Causal Powers versus Regularity-Based Accounts*, in «Memory & Cognition», 23, 243 ff.

WOLFF P., BARBEY A.K. 2015. *Causal Reasoning with Forces*, in «Frontiers in Human Neuroscience», 9.

WOLFF P., BARBEY A.K., HAUSKNECHT M. 2010. *For Want of a Nail: How Absences Cause Events*, in «Journal of Experimental Psychology: General», 139, 191 ff.