Causality and deceit: Do androids watch action movies?

Dusko Pavlovic* Temra Pavlovic

Abstract

We seek causes through science, religion, and in everyday life. We get excited when a big rock causes a big splash, and we get scared when it tumbles without a cause. But our causal cognition is usually biased. The *why* is influenced by the *who*. It is influenced by the *self*, and by *others*. We share rituals, we watch action movies, and we influence each other to believe in the same causes. Human mind is packed with subjectivity because shared cognitive biases bring us together. But they also make us vulnerable.

An artificial mind is deemed to be more objective than the human mind. After many years of science-fiction fantasies about even-minded androids, they are now sold as personal or expert assistants, as brand advocates, as policy or candidate supporters, as network influencers. Artificial agents have been stunningly successful in disseminating artificial causal beliefs among humans. As malicious artificial agents continue to manipulate human cognitive biases, and deceive human communities into ostensive but expansive causal illusions, the hope for defending us has been vested into developing benevolent artificial agents, tasked with preventing and mitigating cognitive distortions inflicted upon us by their malicious cousins. Can the distortions of human causal cognition be corrected on a more solid foundation of artificial causal cognition?

In the present paper, we study a simple model of causal cognition, viewed as a quest for causal models. We show that, under very mild and hard to avoid assumptions, there are always self-confirming causal models, which perpetrate self-deception, and seem to preclude a royal road to objectivity.

*Partially supported by NSF and AFOSR.

Contents

1 Introduction: Causal cognition and its vulnerabilities

1.1 Causal cognition in life and science

Causal cognition drives the interactions of organisms and organizations with their environments at many levels. At the level of perception, spatial and temporal correlations are often interpreted as causations, and directly used for predictions [27]. At the level of science, testing causal hypotheses is an important part of the practice of experiment design [16], although the general concept of causation seldom explicitly occurs in scientific theories [41]. It does occur, of course, quite explicitly in metaphysics and in theory of science [7], and most recently in AI[1][36, 46].

1.2 Causal cognition and launching effects

Causal theories may or may not permeate science (depending on how you look at it) but they certainly permeate life. Why did the ball bounce? Why did my program crash? Why did the boat explode? Who caused the accident? Why did the chicken cross the street? We need to know. We seek to debug programs, to prevent accidents, and to understand chicken behaviors. We often refuse to move on without an explanation. Lions move on, mice run away, but humans want to understand, to predict, to control.

In some cases, the path to a causal hypothesis and the method to test it are straightforward. When a program crashes, the programmer follows its execution path. The cause of the crash must be on it. The causal hypothesis is tested by eliminating the cause and verifying that the effect has been eliminated as well. If a ball bounces like in Newton's cradle, in Fig. 1 on the left, we see that the ball was launched by another ball. If a boat explodes like in Fig. 1 on the right, we see that the explosion was caused by another boat crashing. Action movies are chains of causes and effects, packed between a problem and a solution, which are usually presented as a big cause and a big effect.

In other cases, establishing a causal hypothesis may be hard. It is obvious that the collision caused the explosion; but who caused the collision? It is obvious that the bouncing ball extends the trajectory of the hitting ball; but how is the force transferred from the hitting ball to the bouncing ball through all the balls in-between, that don't budge at all?

Such questions drive our quest for causes, and our fascination with illusions. They drive us into science, and into magic. In a physics lab, a physics teacher would explain the physical

[1] *AI* is usually interpreted as the acronym of *Artificial Intelligence*. But the discipline given that name in 1956 by John McCarthy evolved in the meantime in many directions, some of which are hard to relate with any of the usual meanings of the word "intelligence". We are thus forced to either keep updating the concept of "intelligence" to match the artificial versions, or to unlink the term "AI" from its etymology, and to use it as a word rather than as an acronym, like it happened with the words gif, captcha, gulag, or snafu. The latter choice seems preferable, at least in the present context.

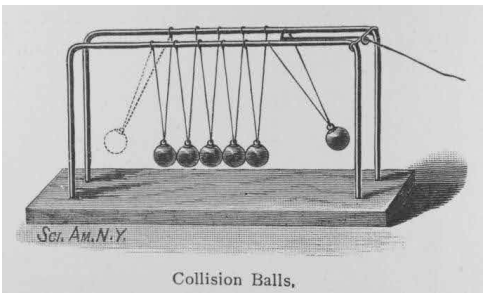


Figure 1: Launching effects

law behind Newton's cradle. But in a magic lab, a magic teacher would present a magic cradle, looking exactly the same like Newton's cradle, but behaving differently. One time the hitting ball hits, and the bouncing ball bounces. Another time the hitting ball hits, but the bouncing ball does not bounce. But later, unexpectedly, the bouncing ball bounces all on its own, without the hitting ball hitting. Still later, the bouncing ball bounces again, and the hitting ball hits *after* that. Magic! Gradually you understand that the magic cradle is controlled by the magic teacher: he makes the balls bounce or stick at his will.

Michotte studied such *"launching effects"* in his seminal book [27]. The action movie industry is built upon a wide range of techniques for producing such effects. While the example in Fig. 1 on the right requires pyrotechnic engineering, the example in Fig. 2 on the right is close to Michotte's lab setups. Observing subtle details, we can usually tell apart a real effect

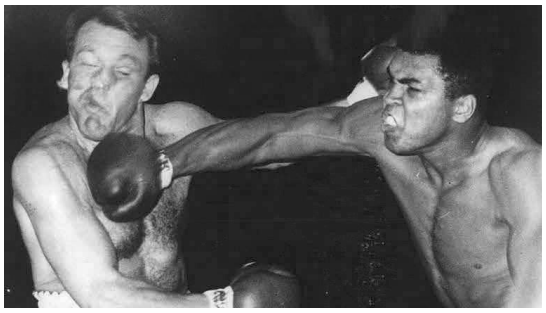


Figure 2: The cause of an effect may appear obvious, but the appearance may deceive

from an illusion. But illusions are often more exciting, or less costly, and we accept them. We enjoy movies, magic, superstition. We abhor randomness and complexity, and seek the simplest causal explanations. We like to be deceived, and we deceive ourselves.

1.3 Causal cognition as a security problem

Cognitive bias and belief perseverance are familiar and well-studied properties of human cognition, promoted in evolution because they strengthen the group cohesion [21]. As human interactions increasingly spread through media, human beliefs are hijacked, their biases are manipulated, and social cohesion is amplified and abused at ever larger scales. This is also well-known and well-studied in psychology [8, Part VI] and in social sciences [49]. However, on the information technology side, cognitive bias manipulations on the web market and in cyberspace have been so profitable in practice, that there was no time for theory.

We live in times of advertising and publicity campaigns. Brands and crowds are built and steered using the same tools, and from the same materials: human cognition and sensitivities. The AI tools, recently added to the social engineering toolkit, appear to be among the most effective ones, and among the least understood. The practice is far ahead of theory. The research reported in this paper is a part of a broader effort [9, 31, 33, 34, 47] towards developing models, methods, and tools for this area, where human and computational behaviors are not just inseparable, but they determine each other; and they do not just determine each other, but they seek to control each other.

The specific motivating question that we pursue is: *Is artificial causal cognition suitable as a tool to defend human causal cognition from the rapid strategic advances in cognitive bias manipulation, influencing, and deceit?*

There are, of course, many aspects of this question, and many approaches to it. The combined backgrounds of this paper's authors provide equipment only for scaling the south-west ridge: where the west cliff of *computation* meets the south cliff of *communication*. Neither of us is equipped to place or interpret our model and results in the context of the extant research in psychology, or social sciences, where they should be empirically validated. We are keen to present it to a mixed audience hoping that there are gates between the silos. Rather than attempt to shed light on the problem from a direction that is unfamiliar to us, we are hoping to shed light on the problem from a direction that is unfamiliar to the reader, while doing our honest best to avoid throwing any avoidable shadows. If the reader sheds some light from their direction, this multi-dimensional problem may emerge in its multi-disciplinary space.

2 Background: Causality theories and models

Causality is studied in different communities from different standpoints. The following high level overview is mainly intended to put us at a starting position. The reader is welcome to fast-forward at any point.

2.1 A very brief history of causality

Causal relations impose order on the events in the world: an event a causes an event b, which causes an event c, whereas the event d is unrelated to a or b, but may contribute to c. An event may have multiple causes, and multiple effects, and they may all be distant from one another in time and in space. We impose order on the world by thinking in terms of causes and effects.

But connecting causes and effects also causes difficulties. In Fig. 1, the force on one end of Newton's cradle causes the effect on the other end without affecting anything in-between; whereas the boats collide and are subsequently engulfed in the explosion, but the explosion is not caused by the collision, but staged. Tellingly, such decouplings are called *special effects*. Our eye distrusts the physical effect on the left, and accepts the special effect on the right. Recognizing such remote causations and adjacent non-causations requires learning about the unobservable causes of observable effects that fill the world, and about the unobservable effects of observable causes, and even about the unobservable causes of unobservable effects. Imposing the causal order on the world forces us to think about the unobservable, whether they are black holes or spirits of our ancestors.

While acting upon some causes to prevent some effects is a matter of individual adaptations and survival for humans and androids, plants and animals, understanding the general process of causation has been a matter of civilization and cognition. While the efforts in the latter direction surely go back to the dawn of mankind, the early accounts go back to pre-Socratic philosophy, and have been traditionally subsumed under *Metaphysics*, which was the (post-Socratic) title of Aristotle's manuscript that came *after* (i.e. it was the *"meta-"* to) his Physics [1]. We mention just three paradigmatic steps towards the concept of causation:

i) Parmenides: "Nothing comes from nothing."
ii) Heraclitus: "Everything has a cause, nothing is its own cause."
iii) Aristotle: "Everything comes from an *Uncaused Cause*."

Step (i) thus introduces the principle of causation; step (ii) precludes causal cycles, and thus introduces the problem of *infinite regression* through causes of causes; and step (iii) resolves this problem by the argument[2] that in Christian theology came to be interpreted as the *cosmological proof* of existence and uniqueness of God. Just like the quest for causes of particular phenomena leads to magic and superstition, the quest for a general Uncaused Cause leads to monotheistic religion. This logical pattern persists in modern cosmology, where

[2] "It is clear, then, that though there may be countless instances of the perishing of unmoved movers, and though many things that move themselves perish and are succeeded by others that come into being, and though one thing that is unmoved moves one thing while another moves another, nevertheless there is something that comprehends them all, and that as something apart from each one of them, and this it is that is the cause of the fact that some things are and others are not and of the continuous process of change; and this causes the motion of the other movers, while they are the causes of the motion of other things. Motion, then, being eternal, the First Mover, if there is but one, will be eternal also; if there are more than one, there will be a plurality of such eternal movers." [1, 258b–259a]

the Uncaused Cause arises as the initial gravitational singularity, known as the Big Bang. Although it is now described mathematically, it still spawns untestable theories. The singularity can be avoided using mathematical devices, inflating and smoothening the Uncaused Cause into a field; but preventing its untestable implications requires logical devices.

While modern views of a *global* cause of the world may still appear somewhat antique, the modern theories of *local* causations evolved with science. The seed of the modern idea that causes can be separated and made testable through *intervention* [18, 45, 48] was sown by Galileo: "That and no other is to be called cause, at the presence of which the effect always follows, and at whose removal the effect disappears" (cited in [7]). As science progressed, philosophers echoed the same idea in their own words, e.g., "It is necessary to our using the word cause that we should believe not only that the antecedent always has been followed by the consequent, but that as long as the present constitution of things endures, it always will be so" [28].

But David Hume remained unconvinced: "When I see, for instance, a billiard-ball moving in a straight line towards another; even suppose motion in the second ball should by accident be suggested to me, as the result of their contact or impulse; may I not conceive, that a hundred different events might as well follow from the cause? [...] The mind can never possibly find the effect in the supposed cause, by the most accurate scrutiny and examination. For the effect is different from the cause, and can never be discovered in it." [22, 4.10] Hume's objections famously shook Immanuel Kant from his metaphysical "dogmatic slumber", and they led him on the path of *"transcendental deduction"* of effects from causes, based on *synthetic a priori* judgements, that are to be found mostly in mathematics [24]. Scientists, however never managed to prove any of their causal hypotheses by pure reason, but took over the world by disproving their hypotheses in lab experiments, whereupon they would always proceed to formulate better causal hypotheses, which they would then try to disprove again, and so on. The imperative of falsifiable hypotheses, and the idea of progress by disproving them empirically, suggested the end of metaphysics, and of causality, and it prompted Bertrand Russell to announce in 1912: "We shall have to give up hope of finding causal laws such as Mill contemplated, as any causal sequence which we have observed may at any moment be falsified without a falsification of any laws of the kind that the actual sciences aim at establishing. [...] All philosophers of every school imagine that causation is one of the fundamental axioms or postulates of science, yet, oddly enough, in advanced science such as [general relativity], the word 'cause' never occurs. [...] The law of causality, I believe, like much that passes muster among philosophers, is a relic of a bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm." [41]

2.2 A very high-level view of causal models

In spite of the logical difficulties, the concept of causation persisted. General relativity resolved the problem of action at distance, that hampered Newtonian theory of gravitation, by replacing causal interactions by fields; but the causal cones continued to mushroom through

the spacetime of special relativity, even along the closed timelike curves. In quantum mechanics, causality was one of the main threads in the discussions between Einstein and Bohr [5]. It was also the stepping stone into Bohm's quantum dialectical materialism [2, 3]. Last but not least, causality remains central in the modern axiomatizations of quantum mechanics [10, 11], albeit in a weak form. Nevertheless, even the weakest concept of causality implies that the observables must be objective, in the sense that the outcomes of measurements should not depend subjective choices, which precludes covert signaling.

The most influential modern theory of causation arose from Judea Pearl's critique of subjective probability, which led him to *Bayesian networks* [35, 37]. As a mathematical structure, a Bayesian net can be thought of as an extension of a Markov chain, where each state is assigned a random variable, and each transition represents a stochastic dependency. Extending Markov chains in this way, of course, induces a completely different interpretation, which is stable only if there are no cycles of state transitions. Hence the causal ordering.

Originally construed as an AI model, Bayesian networks turned out to be a useful analysis tool in many sciences, from psychology [19] to genetics and oncology [42]. Broad overviews of the theory and of the immense range of its applications can be found in [25, 36, 46]. Very readable popular accounts are [38, 39], and [44] is written from the standpoint of psychology. The causal models imposing the acyclicity requirement on the dependencies between random variables opened opened an alley towards an algorithmic approach to causal reasoning, based on dependency discovery: given some input random variables and some output random variables, specify a network of unobservable random variables whereby the given inputs can cause the given outputs. A good introduction into such discovery algorithms is [46]. The applications have had a tremendous impact.

A skeptic could, or course, argue that inserting in-between a cause and an effect a causal diagram is just as justified as inserting Kant's synthetic *a priori* judgements. A control theorist could also argue that the canonical language for specifying dependencies between random variables was provided by stochastic calculus a long time ago. They were not presented as elegantly as in bayesian nets, but they were calculated and graphed. On the other hand, the dependencies modeled in control theory are not required to be acyclic. This is, of course, essential, since the acyclicity requirement precludes feedback. Is feedback acausal? A simple physical system like centrifugal governor[3] obviously obeys the same laws of causation like its feedback-free cousins: an increase in angular momentum causes the valve lever to rise; a wider opening lets more steam out and decreases the pressure; a lower pressure causes a decrease in angular momentum. The only quirk is thus that some physical effects thus cause physical effects on their causes, in continuous time. Whether that is causation or not depends on the modeling framework.

Be it as it may, our very short story about the concepts and the models of causality ends

[3]Centrifugal governor consists of a pair of rotating weights, mounted to open a valve proportionally to their rotation speed, were used to control pressure in steam engines, and before that on the windmill stones. Bayesian nets require an added dimension of time to model such things, basically unfolding the feedback!

here. The upshot is that ***there are effective languages for explaining causation***. The availability and effectiveness of such languages is the central **assumption** on which the rest of the paper is based. Such languages, effective enough to explain everything, have been provided by metaphysics, and by transcendental deduction, and by bayesian networks, and by stochastic processes. There are other such languages, and there will be more. The rest of our story proceeds the same for any of them.

3 Approach: Causes in boxes, tied with strings

3.1 String diagrams

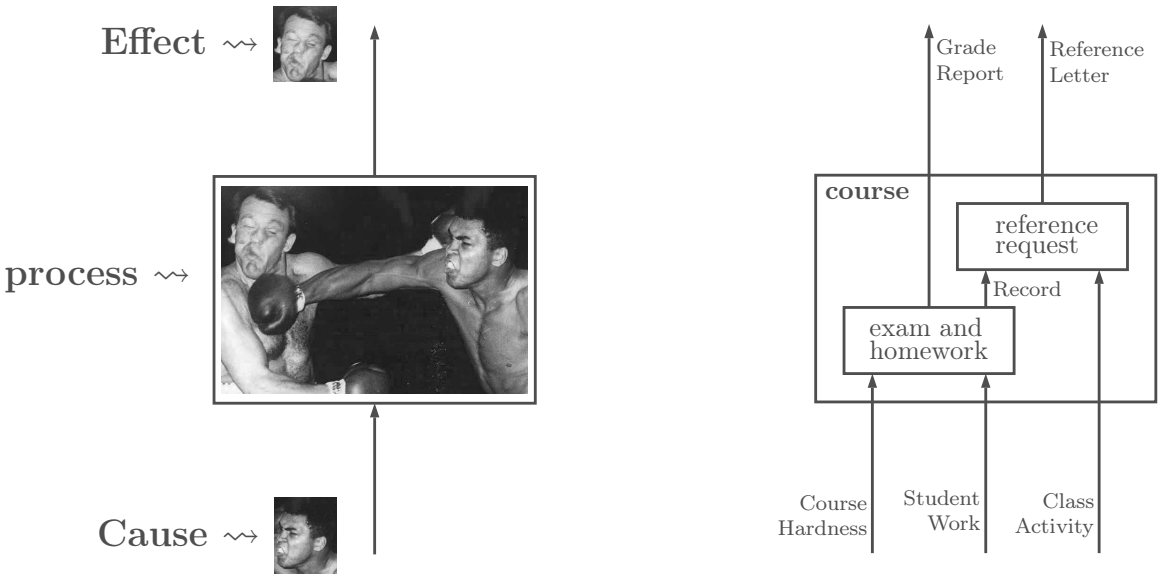


Figure 3: String diagrams: causation flows upwards

Henceforth, we zoom out, hide the "innards" of causal models, and study how they are composed and decomposed. Towards this goal, causal processes are presented as *string diagrams*, like in Fig. 3. String diagrams consist of just two graphic components:

- ***strings*** — representing *types*, and
- ***boxes*** — representing causal *processes*.

Formally, types and processes are the basic building blocks of a causal modeling language: e.g., they may correspond to random variables and stochastic processes. Informally, a type can be thought of as a collection of events. In a causal process, enclosed in a box, the events of the input type, corresponding to the string that enters the box at the bottom, cause the events of the output type, corresponding to the string that exits the box at the top. The time just flows upwards in our string diagrams. We call the types consumed and produced

by a causal process inputs and outputs (and not causes and effects) to avoid the ambiguities arising from the fact that the events produced by one process as effects may be consumed by another process as causes. The input and the output type may be the same. The diagram in Fig. 3 on the left is hoped to convey an idea what is what. Presenting causal processes as boxes allows us to abstract away the details when we need a high-level view, and also to refine the view as needed, by opening some boxes and displaying more details. This is similar to the mechanism of virtual functions in programming: we specify the input and the output types, but postpone the implementation details. A refined view of a causal process in an open box may be composed of other boxes connected by strings. An example[4] is in Fig. 3 on the right. The types of the strings coming in and out show that *grades* and *reference letters*, as events produced as effects of a causal process **course**, are causally determined by *students' work* and *class activities*, as well as by the *hardness* of the course itself. All these causal factors are viewed of events of suitable types. When we open the **course** box, we see that this causal process is composed from two simpler causal processes: one is **exam and homework**, the other **reference request**. Each of them inputs two causal factors at the bottom; **exam and homework** outputs two effects, whereas **reference request** outputs one. Each *grade*, as an event of type Grade Report, is caused both by the *course hardness* and by the *student work*; whereas the *reference letters* are caused (influenced, determined...) by the *class activity* and by the course *record*, which is again an effect of the *student work* and of the *course hardness* in the process of **exam and homework**. The causal dependencies between the random variables corresponding to the types Grade Report and Reference Letter, and their dependencies on the random variables corresponding to Course Hardness, Student Work, and Class Activity, are thus displayed in the string diagram inside the **course** box. For a still more detailed view of causal relations, we could zoom in further, and open the boxes **exam and homework** and **reference request**, to display the dependencies through some other random variables, influenced by some other causal processes.

The causes corresponding to the input types are assumed to be independent on each other. More precisely, the random variables, corresponding to the types Course Hardness, Student Work, and Class Activity, are assumed to be statistically independent. On the other hand, the causal dependencies of random variables corresponding to Grade Report, Record, and Reference Letter are displayed in the diagram. E.g., the content of a Reference Letter is not directly caused by Course Hardness, but it indirectly depends on it, since the student performance in the Record depends on it, and the Record is taken into account in the Reference Letter.

The abstract view of the causal process **box match** on the left could be refined in a similar way. The causes are boxers' actions, the effects are boxers' states. The images display a particular cause, and a particular effect of these types. The direct cause of the particular effect that is displayed on top is a particular blow, that is only visible inside the box. The cause at the bottom is the boxer's decision to deliver that particular blow. The causal process transforming this causal decision into its effect can be refined all the way to distinguishing

[4]This has been a running example in causal modeling textbooks and lectures at least since [25].

good boxing from bad boxing, and to analyzing the causes of winning or losing.

Spelling out a process corresponding to the **box match** on the right in Fig. 2 might be even more amusing. While the strings outside the box would be the same, the causal dependencies inside the box would be different, as boxers' states are not caused by their blows, but feigned; and the blows are not caused by boxers' decisions, but by the movie director's requests.

3.2 A category of causal processes

String diagrams described in the previous section provide a graphic interface for any of causal modeling languages, irrespective of their details, displaying only how causal processes are composed. It is often convenient to arrange compositional structures into *categories* [26]. The categorical structures capturing causations turn out to yield to string diagrams [13, 12, 23]. For the most popular models, as mentioned above, the strings correspond to random variables, the boxes to parametrized stochastic processes. The strings are thus the objects, the boxes the morphisms of a monoidal category. When no further constraints are imposed, this category can be viewed as the coslice, taken from the one-point space of the category of free algebras for a suitable stochastic monad. For a categorical account of full stochastic calculus, the monad can be taken over measurable spaces [15, 17]. For a simplified account the essential structures, the convexity monad over sets may suffice [23]. For our purpose of capturing causal explanations *and* taking them into account as causal factors, we must deviate from the extant frameworks [13, 15, 17, 23][5] in two ways. One is minor, and well tried out: the randomness is captured by *sub*probability measures, like in [29], to account for nontermination. The other is that causal processes must be taken up to *stochastic indistinguishability*, for the reasons explained in [46, Ch. 4].

Notation. In the rest of the paper, we fix an abstract universe $\mathcal{C}$ of causal processes, which happens to be a strict monoidal category [26, Sec. VII.1]. This structure is precisely what the string diagrams display. The categorical notation collects all strings together, in the family of event *types* $|\mathcal{C}|$; and for any pair of types $A, B \in |\mathcal{C}|$ it collects all boxes together, in the family $\mathcal{C}(A, B)$ of causal *processes*. In summary,

- ***strings*** form in the family of *types*

$$|\mathcal{C}| \;=\; \{A, B, C \ldots, \text{Student Work}, \text{Grade Report} \ldots\}$$

- ***boxes*** form for each pair $A, B \in |\mathcal{C}|$ the family of *processes*,

$$\mathcal{C}(A, B) \;=\; \{f, g, t, u, \ldots, \textbf{box match}, \textbf{course} \ldots\}$$

The compositional structure is briefly described in the Appendix.

[5] There is an entire research area of categorical probability theory, steadily built up since at least the 1960s, with hundreds of references. The three cited papers are a tiny random sample, which an interested reader could use as a springboard into further references, in many different flavors.

4 Modeling: Causal models as causal factors

4.1 Modeling causal models

The experimenter has been recognized as a causal factor in experiments since the onset of science. Already Galileo's concept of causality, mentioned in Sec. 2, requires that the experimenter manipulates a cause in order to establish how the effect depends on it. In modern theories of causation, this idea is refined to the concept of *intervention* [20]. In quantum mechanics, the effects of measurements depend not only on the causal freedom of the measured particles, but also on experimenters' own causal freedom [4, 14]. As mentioned in Sec. 1, similar questions arise even in magic: if the magician manipulates the effects, what causes magician's manipulations?

But if the magician and the experimenter are causal processes themselves, then they also occur, as boxes among boxes, in the universe $\mathcal{C}$ of causal processes. And if the experimenter's causal hypotheses are caused by some causal factors themselves, then the universe $\mathcal{C}$ contains, among other types, a distinguished type of causal models Ω, where the experimenter outputs his hypotheses. For instance, Ω can be the type of our string diagram models, like in Fig. 3. Or Ω can be the type of (suitably restricted) stochastic processes; or of Bayesian nets, including the running example [25, Fig. 3.3], on which Fig. 3 is based.[6] There are many different languages for describing different causations, and thus many different ways to think of the type Ω. Psychologist's descriptions of causal cognition differ from physicist's view of the causal process of experimentation. What is their common denominator? *What distinguishes Ω from other types?*

In the Sec. 4.3, we attempt to answer this question by specifying a structure, using three simple axioms, which should be carried by the type Ω in all frameworks for causal modeling. The other way around, we take this structure to mean that a framework where it occurs describes causal modeling; that the blind men talking about different parts of an elephant are talking about the same animal. This is where the **assumption** stated at the end of Sec. 2 is used.

But stating and using the axioms requires two basic syntactic features, which are explained next.

4.2 Parametrizing and steering models and processes

Suppose that we want to explore how learning environments influence the causal process **course** from Fig. 3. How does the causal impact of Course Hardness change depending on

[6] One of the coauthors of this paper thinks of Ω as typing the well-formed expressions in any of the suitable modeling languages, discussed in Sec. 2.2. The other coauthor is inclined to include a wider gamut of causal theories, including those touched upon in Sec. 2.1. Different theories may not only describe different models, but also prescribe different interpretations.

the *school* where a **course** takes place? Abbreviating for convenience the cause and the effect types of the causal process **course** to

$$\begin{aligned} \text{Causes} &= \text{Course Hardness} \otimes \text{Student Work} \otimes \text{Class Activity} \\ \text{Effects} &= \text{Grade Report} \otimes \text{Reference Letter} \end{aligned}$$

we now have

- a family of causal processes Causes $\xrightarrow{\textbf{course}\,(shool)}$ Effects, indexed over $school \in$ Schools, or equivalently
- a parametric causal process Schools $\otimes$ Causes $\xrightarrow{\textbf{course}}$ Effects.

In the first case, each causal process **course***(school)* consumes the causal inputs in the form **course***(school)(hardness, work, activity)*; whereas in the second case, all inputs are consumed together, in the form **course***(school, hardness, work, activity)*. The difference between the parameter *school* and the general causal factors $hardness, work, activity$ is that the parameter is *deterministic*, whereas the general causal factors influence processes with some *uncertainty*. This means that entering the same parameter *school* always produces the same causal process, whereas the same causal factors $hardness, work, activity$ may have different effects in different samples.

The convenience of internalizing the indices and writing indexed families as parametric processes is that steering processes can be internalized as reparametrizing. E.g. given a function ς : Teachers $\longrightarrow$ Schools, mapping each *teacher* the the *school* where they work, induces the reindexing of families

$$\text{Causes} \xrightarrow{\textbf{course}\,(shool)} \text{Effects} \qquad school \in \text{Schools}$$

$$\text{Causes} \xrightarrow{\textbf{course}\,(\varsigma\,(teacher))} \text{Effects} \qquad teacher \in \text{Teachers}$$

which can however be captured as a single causal process in the universe $\mathcal{C}$

$$\text{Teachers} \otimes \text{Causes} \xrightarrow{\varsigma \,\otimes\, \text{Causes}} \text{Schools} \otimes \text{Causes} \xrightarrow{\textbf{course}} \text{Effects}$$

where the causal factor s happens to be deterministic. In general, an arbitrary causal process $Y \otimes A \xrightarrow{\textbf{process}} B$ can be steered along an arbitrary function $\texttt{steer} : X \longrightarrow Y$, viewed as a deterministic process:

$$X \otimes A \xrightarrow{\texttt{steer} \otimes \texttt{A}} Y \otimes A \xrightarrow{\textbf{process}} B$$

Deterministic functional dependencies are characterized in string diagrams in Appendix D.

4.3 Axioms of causal cognition

Every framework for causal modeling must satisfy the followin axioms:

I : Every causal model models a unique causal process (over the same parameters).
II : Every causal process has a model (not necessarily unique).
III : Models are preserved under steering.

4.3.1 Axioms formally

We view a causal model as a family $P(y) \in \Omega$, indexed by $y \in Y$, i.e. as a parametrized model $Y \xrightarrow{P} \Omega$. The notation introduced in Sec. 3.2 collects all Y-parametrized causal models in the set of processes $\mathcal{C}(Y, \Omega)$. The parameter type Y is arbitrary. On the other hand, all Y-parametrized causal processes $Y \otimes A \xrightarrow{p} B$, where the events of type A cause events of type B, are collected in the set $\mathcal{C}(Y \otimes A, B)$.

Axiom I postulates that every parametrized causal model $Y \xrightarrow{P} \Omega$ induces a unique causal process $Y \otimes A \xrightarrow{[\![P]\!]} B$ with the same parameters Y. A causal modeling framework is thus given by a family of the *prediction* maps

$$\mathcal{C}(Y, \Omega) \quad \xrightarrow{[\![-]\!]} \quad \mathcal{C}(Y \otimes A, B) \tag{1}$$

indexed over all types Y, A and B.

Axiom II says that the prediction maps $[\![-]\!]$ are surjective: for every causal process $Y \otimes A \xrightarrow{p} B$ there is a causal model $Y \xrightarrow{P} \Omega$ that predicts its behavior, in the sense that the process $[\![P]\!]$ is *indistinguishable* from p, which we write $[\![P]\!] \approx p$. The indistinguishability relation $\approx$ is explained in the next section.

Axiom III says that for any function $\varsigma : X \longrightarrow Y$ and any causal model $Y \xrightarrow{P} \Omega$ reparametrizing P along ς models steering the modeled process $[\![P]\!]$ along it, in the sense that

$$[\![X \xrightarrow{\varsigma} Y \xrightarrow{P} \Omega]\!] \quad = \quad X \otimes A \xrightarrow{\varsigma \otimes A} Y \otimes A \xrightarrow{[\![P]\!]} B \tag{2}$$

4.3.2 Axioms informally

Axiom I is a soundness requirement: it says that every causal model models some causal process. If causal processes are viewed as observations, the axiom thus says that for any causal model, we will recognize the modeled process if we observe it.

Axiom II is a completeness requirement: it says that for every causal process that we may observe, there is a causal model that models it. Can we really model everything that we observe? Yes, but our models are valid, i.e. their predictions are consistent with the behavior of the modeled processes *only in so far as our current observations go*. Given a process p, we can always find a model P whose predictions $[\![P]\!]$ summarize our observations of p, so that $[\![P]\!]$ and p are for us *indistinguishable*, i.e. $[\![P]\!] \approx p$. The less we observe, the less we distinguish, the easier we model.

Indistinguishability relations, be it statistical, computational, or observational, are central in experimental design, in theory of computation, and in modern theories of causation

[46, Ch. 4]. The problem of distinguishing between observed processes has been tackled since the early days of statistics by significance testing, and since the early days of computation by various semantical and testing equivalences. Axiom II says that, up to an indistinguishability relation, any observed causal process has a model. It does not say anything about the hardness of modeling. This problem is tackled in research towards *cause discovery algorithms* [46].

Axiom III is a coherence requirement: it says that steering a process $Y \otimes A \xrightarrow{p} B$ along a deterministic function $\varsigma : X \longrightarrow Y$ does not change the causation, in the sense of (2) or

$$[\![P]\!] \approx p \quad \Longrightarrow \quad [\![P \circ \varsigma]\!] \approx \big(p \circ (\varsigma \otimes A)\big) \tag{3}$$

4.4 Universal testing

Since any causal modeling language can thus be viewed as a type Ω, living in the process universe $\mathcal{C}$, the family of all causal models $\omega \in \Omega$, trivially indexed over itself, can be represented by the identity function $\Omega \xrightarrow{Id} \Omega$, viewed as a Ω-parametrized model. Instantiating in (2) Id for P (and thus Ω for Y) yields $[\![\varsigma]\!] = [\![Id]\!] \circ (\varsigma \otimes A)$, for any $X \xrightarrow{\varsigma} \Omega$. *Mutatis*

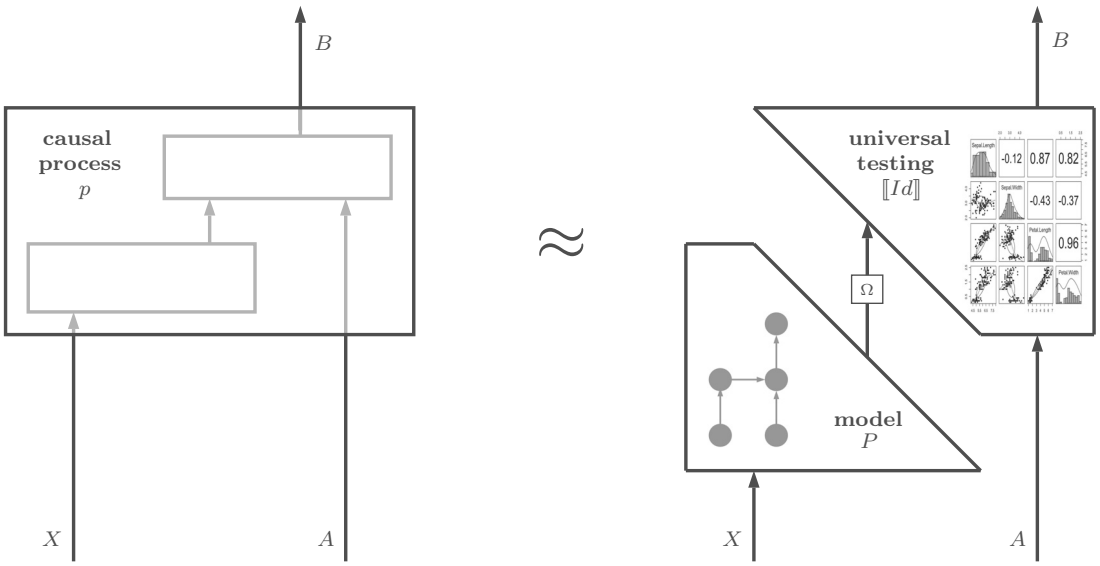


Figure 4: There is an explanation for every causation

mutandis, for an arbitrary process p and a model P assured by axioms I and II, axiom II thus implies $p \approx [\![Id]\!] \circ (P \otimes A)$, as displayed in Fig. 4.

Any causal universe $\mathcal{C}$, as soon as it satisfies axioms I–III, thus contains, for any pair A, B, a ***universal testing*** process $\Omega \otimes A \xrightarrow{[\![Id]\!]} B$, which inputs causal models and tests their

predictions, in the sense that $X \otimes A \xrightarrow{P \otimes A} \Omega \otimes A \xrightarrow{[\![Id]\!]} B$ is indistinguishable from $X \otimes A \xrightarrow{p} B$ whenever $[\![P]\!]$ is indistinguishable from p. Universal testing is thus a causal process where the predictions of causal models are derived as their effects. It can thus be construed as a very high level view of scientific practice; perhaps also as an aspect of cognition.

4.5 Partial modeling

If a causal process has multiple causal factors, then they can be modeled separately be treating some of them as model parameters. E.g. a process in the form $Y \otimes X \otimes A \xrightarrow{r} B$ can be viewed as a Y-parametrized process with causal factors of type $X \otimes A$, or as a $Y \otimes X$-parametrized process with causal factors of type A. The two different instances of (1), both surjections by axiom II, would lead to models $Y \xrightarrow{R'} \Omega$, and $Y \otimes X \xrightarrow{R''} \Omega$, with $r \approx [\![R']\!]_Y \approx [\![R'']\!]_{Y \otimes X}$ by (1), and thus $r \approx [\![Id]\!]_{X \otimes A} \circ (R' \otimes X \otimes A) \approx [\![Id]\!]_A \circ (R'' \otimes A)$ by Fig. 4.

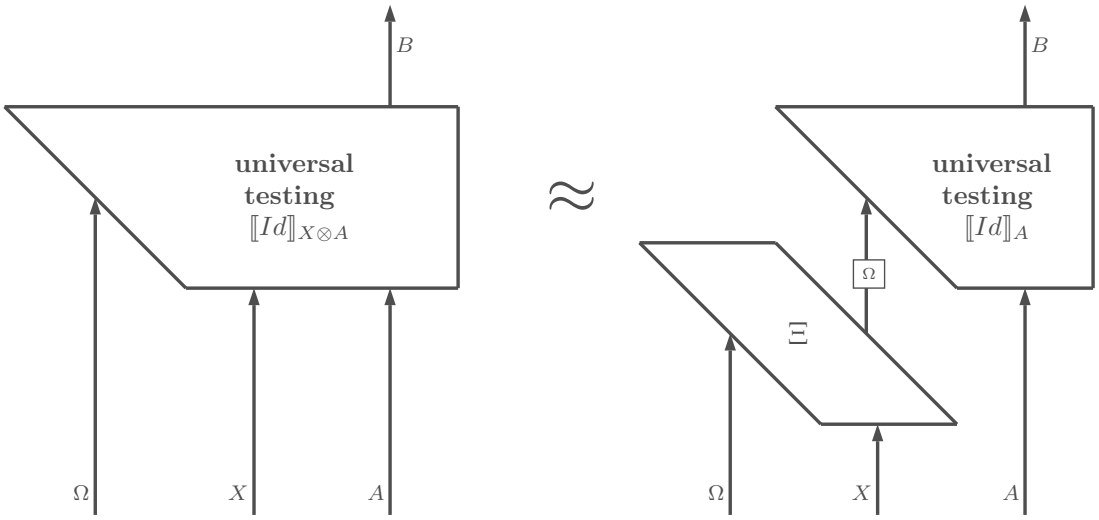


Figure 5: Causes of type X are treated as parameters

In particular, taking r to be the universal process $\Omega \otimes X \otimes A \xrightarrow{[\![Id]\!]} B$, and interpreting it as an $\Omega \otimes X$-parametrized process, leads to an $\Omega \otimes X$-parametrized model Ξ, such that $[\![Id]\!]_{X \otimes A} \approx [\![Id]\!]_A \circ (\Xi \otimes A)$, as displayed in Fig. 5.

4.6 Slicing models

Using universal testing processes and partial modeling, causal processes can be modeled incrementally, factor by factor, like in Fig. 6.

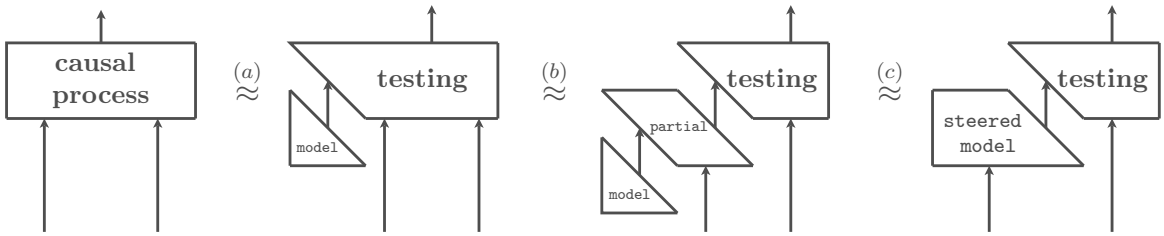


Figure 6: Separating independent causes allows incremental modeling

5 Construction: Self-confirming causal models

In everyday life, we often defend our views by steering ourselves to the standpoints where they seem valid. This assures belief perseverance, but also leads to cognitive bias and self-deception. In science, experimenters obviously influence the outcomes of their experiments by choosing what to measure. In quantum mechanics, this influence propagates to processes themselves, as the choice of the measurement basis determines the choice of the states to which the measured process may collapse [5, 14]. Causal models are thus always among the causal factors of the modeled quantum processes[7]. Formally, a quantum process is thus always in the form $\Omega \otimes A \xrightarrow{q} B$, parametrized by its models in Ω.

Many processes outside the realm of quantum are in this form as well, as causal cognition curls into itself by influencing causation. The placebo and the nocebo effects are ubiquitous: patient's belief in the effectiveness of a medication contributes to its effectiveness, whereas negative beliefs often cause negative effects. In Shakespeare's tragedy, Macbeth is driven to murder the King by the prophecy that he would murder the King. In the early days of Facebook, attracting new members required convincing them that many of their friends were already members. This initially had to be a lie, but many believed it, joined, and it ceased to be a lie.

Other causal models impact their own validity negatively. If the customers of a restaurant come to believe that it is too busy on Fridays, it may end up empty on Fridays: the belief will invalidate itself. If the rumor that it was empty spreads, the customers will swarm back, and the model will invalidate itself again. Such network effects are observed not only in the famous El Farol bar in Santa Fe, and in Kolkatta paise restaurants, but also in financial and stock markets, and in urban agglomerations. Causal models often influence the modeled causal structures, and sometimes impact their own validity. — *What are the conditions and limitations of this phenomenon?*

Fig. 7 formalizes this question. Under which conditions can a process $\Omega \otimes A \xrightarrow{q} B$ be steered by a model $\Gamma \in \Omega$ to a process which confirms Γ's predictions, i.e. such that $q \circ (\Gamma \otimes A) \approx [\![\Gamma]\!]$?

[7] "The complete freedom of the procedure in experiments common to all investigations of physical phenomena, is in itself of course contained in our free choice of the experimental arrangement." [6]

By reinterpreting the outputs of q, under certain additional conditions, the same schema can

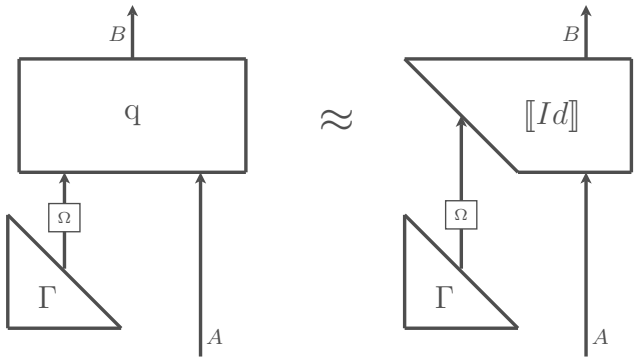


Figure 7: The process q steered by a self-confirming model Γ into $q(\Gamma, a) \approx [\![\Gamma]\!](a)$

be used to construct models that would invalidate themselves.

In any case, it turns out that a causal universe $\mathcal{C}$ *contains a self-confirming causal model for every causal process parametrized over models*, as soon as it supports axioms I–III.

The crucial insight is that the partial model $\Xi : \Omega \otimes \Omega \longrightarrow \Omega$ from Sec. 4.5 induces the Ω-parametrized model $\Omega \xrightarrow{\Delta} \Omega \otimes \Omega \xrightarrow{\Xi} \Omega$, or in the indexed form $\Xi(\omega, \omega)$, which for every $\omega \in \Omega$ predicts the effect of steering the Ω-parametrized process $[\![\omega]\!]$ to ω. This "self-model" $\Xi \circ \Delta$ is displayed in Fig. 8 on the left, where we precompose the given process q with it. Axiom

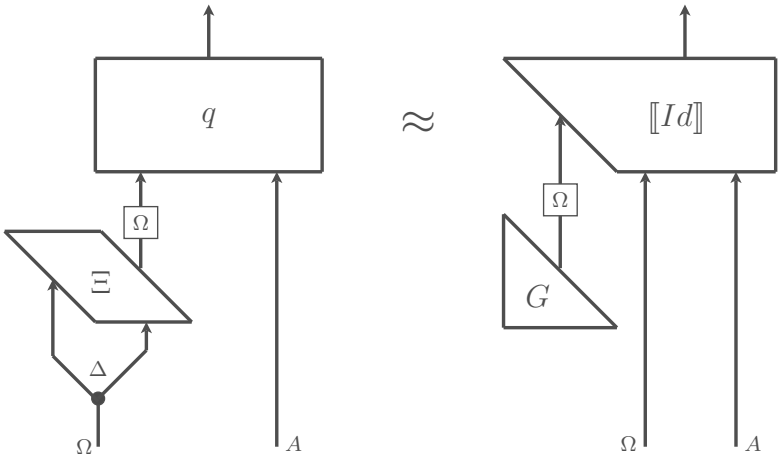


Figure 8: Modeling self-modeling

II now gives a model G of the resulting composite process. Fig. 9 shows that $\Gamma = \Xi \circ \Delta \circ G$ provides the claimed causal model of the given process $\Omega \otimes A \xrightarrow{q} B$, parametrized over the possible models. The first step follows from the definition of G in Fig. 8, by substituting G also as the input. The second step folds two copies of G into one. The third step follows

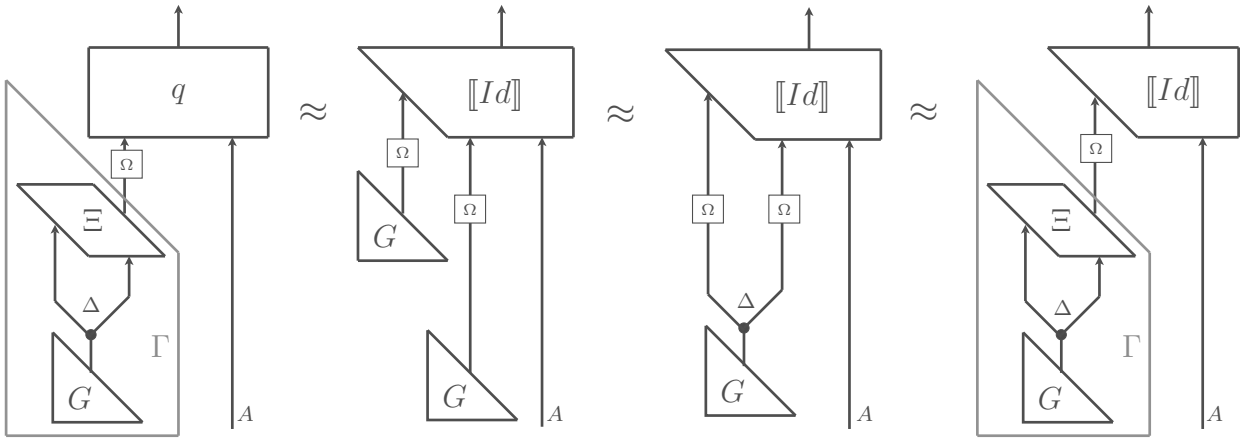


Figure 9: Setting $\Gamma = \Xi \circ \Delta \circ G$ gives $q \circ (\Gamma \otimes A) \approx [\![\Gamma]\!]$

from the definition of Ξ in Fig. 5. The upshot is that the causal process $[\![Id]\!] \circ (\Gamma \otimes A)$ on the right, which is by Sec. 4.4 just the process $[\![\Gamma]\!]$ modeled by Γ, is indistinguishable from the instance $q \circ (\Gamma \otimes A)$ on the left, obtained by steering the process q, which is parametrized over its models, to the model Γ. The causal model Γ thus steers the given causal process q to an instance which confirms Γ as its model.

6 Summary: Towards artificial causal cognition?

The concept of causality made a remarkable historic voyage, from ancient physics and metaphysics [1] to modern physics [2, 10, 11], through incisive critiques in philosophy [22, 41], through enduring attention in psychology [27, 18, 44], all the way to its current promise in machine learning and in AI [25, 36, 37, 39, 40, 46]. How much does the new outlook change the scene of causal cognition?

If subjectivity is implied by cognitive self-confirmation, then the results presented in this paper suggest that all causal cognition, human or artificial, is inexorably subjective, as soon as models cannot be eliminated as causal factors. In particular, we have shown how causal models and causal processes can be steered to support each other, and adapted to absorb any new evidence. This is a concerning conclusion, if scientific theories are required to be falsifiable. The fact that causal cognition, as modeled here, permits steering and self-validation under the assumptions as mild as axioms I–III, can also be interpreted as demonstrating that causal cognition cannot be understood just in terms of logic and structure. Such a conclusion would have significant repercussions on the AI research.

Computer science has been recognized as a natural science almost from the outset.This is by now largely accepted in principle, although often misunderstood in the popular narrative, and dismissed as irrelevant by many working computer scientists and engineers. With the

advent of the web, and the spread of personal devices, computation propagated through the vast area of economic, political, and social processes, and came to play a role in many forms of human cognition. Changing the angle only slightly, the same process can be viewed as evolution of artificial cognition within a natural process of network computation. Whether the androids of the future will be attracted to causal illusions of action movies remains to be seen; but the humanities of the present cannot avoid studying human cognition together with artificial cognition. Like identical twins, or like parallel universes, or like Lao Tse and the butterfly, the two can be reliably distinguished only by themselves, observing each other from within, but difficult to tell apart for the external observers. The human gave rise to the artificial, and then the artificial gave rise to a new human, and then maybe there was a new artificial, and then they lost count.

Psychology of artificial mind? Herbert Simon, one of the originators AI, anticipated the creation of a whole new realm of *"The Sciences of the Artificial"* [43]. With the AI technologies permeating market and politics, a *psychology of the artificial* seems to have become not just a real possibility, but also an urgent task.

References

[1] Aristotle. *Physics*. Clarendon Aristotle series. Clarendon Press, 1983.

[2] David Bohm. *Causality and Chance in Modern Physics*. Taylor & Francis, 2004.

[3] David Bohm and Chris Talbot (ed.). *David Bohm: Causality and Chance, Letters to Three Women*. Springer International Publishing, 2017.

[4] Niels Bohr. Causality and complementarity. *Philosophy of Science*, 4(3):289–298, 1937.

[5] Niels Bohr. On the notions of causality and complementarity. *Science*, 111(2873):51–54, 1950.

[6] Niels Bohr. The causality problem in atomic physics. In Jürgen Kalckar, editor, *Foundations of Quantum Physics II (1933–1958)*, volume 7 of *Niels Bohr Collected Works*, pages 299 – 322. Elsevier, 1996.

[7] Mario Bunge. *Causality and Modern Science: Third Revised Edition*. Dover Publications, 2012.

[8] David M. Buss. *The Handbook of Evolutionary Psychology, Volume 2: Integrations*. The Handbook of Evolutionary Psychology. Wiley, 2015.

[9] Jason Castiglione, Dusko Pavlovic, and Peter-Michael Seidel. Privacy protocols. In Joshua Guttman, editor, *CathFest: Proceedings of the Symposium in Honor of Catherine Meadows*, volume 11565 of *Lecture Notes in Computer Science*, pages 167–192. Springer, 2019.

[10] Giulio Chiribella, Giacomo Mauro D'Ariano, and Paolo Perinotti. Informational derivation of quantum theory. *Physical Review A*, 84:012311, 2011.

[11] Bob Coecke and Aleks Kissinger. Categorical Quantum Mechanics I: Causal Quantum Processes. In *Categories for the Working Philosopher*, chapter 12, pages 286–328. Oxford University Press, 2017.

[12] Bob Coecke and Aleks Kissinger. *Picturing Quantum Processes*. Cambridge University Press, 2017.

[13] Bob Coecke and Robert W. Spekkens. Picturing classical and quantum Bayesian inference. *Synthese*, 186(3):651–696, 2012.

[14] John Conway and Simon Kochen. The strong free will theorem. *Notices of the AMS*, 56(2):226–232, 2009.

[15] Jared Culbertson and Kirk Sturtz. A categorical foundation for bayesian probability. *Applied Categorical Structures*, 22(4):647–662, 2014.

[16] R.A. Fisher. *The Design of Experiments*. Oliver and Boyd, 1935.

[17] Michèle Giry. A categorical approach to probability theory. In Claudia Casadio et. al., editor, *Categorical aspects of topology and analysis (Ottawa, Ont., 1980)*, volume 915 of *Lecture Notes in Math.*, pages 68–85. Springer Verlag, 1982.

[18] Alison Gopnik and Laura Schulz, editors. *Causal Learning: Psychology, Philosophy, and Computation*. Oxford Series in Cognitive Development. Oxford University Press, 2007.

[19] Alison Gopnik and Joshua B. Tenenbaum. Bayesian networks, bayesian learning and cognitive development. *Developmental Science*, 10(3):281–287, 2007.

[20] York Hagmayer, Steven Sloman, David Lagnado, and Michael R. Waldmann. Causal reasoning through intervention. In Alison Gopnik and Laura Schulz, editors, *Causal Learning: Psychology, Philosophy, and Computation*, chapter 6, pages 86–100. Oxford University Press, Oxford, 2007.

[21] Martie G. Haselton, Daniel Nettle, and Damian R. Murray. The evolution of cognitive bias. In David M. Buss, editor, *The Handbook of Evolutionary Psychology, Volume 2: Integrations*, chapter 41, pages 968–987. Wiley, 2015.

[22] David Hume. *An Enquiry Concerning Human Understanding*. Hackett Classics Series. Hackett Publishing Company, 1993.

[23] Bart Jacobs, Aleks Kissinger, and Fabio Zanasi. Causal inference by string diagram surgery. In Mikolaj Boja'nczyk and Alex Simpson, editors, *Proceedings of FoSSaCS 2019*, volume 11425 of *Lecture Notes in Computer Science*, pages 313–329. Springer, 2019.

[24] Immanuel Kant. *Critique of Pure Reason*. Dover Philosophical Classics. Dover Publications, 2012.

[25] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques*. Adaptive computation and machine learning. MIT Press, 2009.

[26] Saunders Mac Lane. *Categories for the Working Mathematician*. Number 5 in Graduate Texts in Mathematics. Springer-Verlag, 1971.

[27] Albert Michotte. *La perception de la causalité*. Études de philosophie. Editions de l'Institut Supérieur de Philosophie, 1946.

[28] J.S. Mill. *A System of Logic Rationative and Inductive*. Longmans, Green & Company, 1904.

[29] P. Panangaden. *Labelled Markov Processes*. Imperial College Press, 2009.

[30] Dusko Pavlovic. Categorical logic of names and abstraction in action calculus. *Math. Structures in Comp. Sci.*, 7:619–637, 1997.

[31] Dusko Pavlovic. Gaming security by obscurity. In Carrie Gates and Cormac Hearley, editors, *Proceedings of NSPW 2011*, pages 125–140, New York, NY, USA, 2011. ACM. arxiv:1109.5542.

[32] Dusko Pavlovic. Geometry of abstraction in quantum computation. *Proceedings of Symposia in Applied Mathematics*, 71:233–267, 2012. arxiv.org:1006.1010.

[33] Dusko Pavlovic. Towards a science of trust. In *Proceedings of the 2015 Symposium and Bootcamp on the Science of Security*, HotSoS '15, pages 3:1–3:9, New York, NY, USA, 2015. ACM. arxiv.org:1503.03176.

[34] Dusko Pavlovic and Catherine Meadows. Actor Network Procedures. In Ram Ramanujam and Srini Ramaswamy, editors, *Proceedings of International Conference on Distributed Computing and Internet Technologies 2012*, volume 7154 of *Lecture Notes in Computer Science*, pages 7–26. Springer Verlag, 2012. arxiv.org:1106.0706.

[35] Judea Pearl. Bayesian networks: A model of self-activated memory for evidential reasoning. In *Proceedings of Cognitive Science Society (CSS-7)*, 1985.

[36] Judea Pearl. *Causality*. Causality: Models, Reasoning, and Inference. Cambridge University Press, 2009.

[37] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann series in representation and reasoning. Elsevier Science, 2014.

[38] Judea Pearl. The seven tools of causal inference, with reflections on machine learning. *Commun. ACM*, 62(3):54–60, February 2019.

[39] Judea Pearl and Dana Mackenzie. *The Book of Why: The New Science of Cause and Effect*. Basic Books, 2018.

[40] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of Causal Inference: Foundations and Learning Algorithms*. Adaptive Computation and Machine Learning series. MIT Press, 2017.

[41] Bertrand Russell. On the notion of cause. *Proceedings of the Aristotelian Society*, 13:1–26, 1912.

[42] Eran Segal, Nir Friedman, Daphne Koller, and Aviv Regev. A module map showing conditional activity of expression modules in cancer. *Nature Genetics*, 36(10):1090–1098, 2004.

[43] Herbert A. Simon. *The Sciences of the Artificial*. The MIT Press. MIT Press, 1996.

[44] Steven A. Sloman. *Causal Models: How People Think About the World and Its Alternatives*. Oxford University Press, 2005.

[45] Steven A. Sloman and York Hagmayer. The causal psycho-logic of choice. *Trends in Cognitive Sciences*, 10(9):407 – 412, 2006.

[46] Peter Spirtes, Clark Glymour, and Richard Scheines. *Causation, Prediction, and Search*. Adaptive computation and machine learning. MIT Press, 2000.

[47] Vladimir Vovk and Dusko Pavlovic. Universal probability-free prediction. *Ann. Math. Artif. Intell.*, 81(1-2):47–70, 2017. arxiv.org:1603.04283.

[48] James Woodward. *Making Things Happen: A Theory of Causal Explanation*. Oxford Studies in Philosophy of Science. Oxford University Press, 2005.

[49] Shoshana Zuboff. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs, 2019.

Appendix

A Composing and decomposing causal processes

The salient feature of the string diagram presentation of processes is that the two dimensions of string diagrams correspond to two kinds of function composition. This is displayed in Fig. 10.

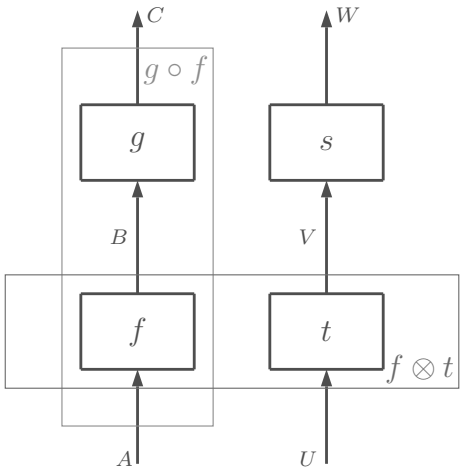


Figure 10: Sequential composition $g \circ f$ and parallel composition $f \otimes t$

- ***Sequential composition*** of causal processes corresponds to linking the corresponding string diagrams *vertically*: the effects of the causal process $A \xrightarrow{f} B$ are passed to the causal process $B \xrightarrow{g} C$ to produce the causal process $A \xrightarrow{g \circ f} C$;
- ***parallel composition*** lays causal processes next to each other *horizontally*: the processes $A \xrightarrow{f} B$ and $U \xrightarrow{t} V$ are kept independent, and their composite $A \otimes U \xrightarrow{f \otimes t} B \otimes V$ takes each processes causal factors separately, and produces their effects without any interactions between the two:

As categorical structures, these operations are captured as the mappings

$$\begin{aligned} \mathcal{C}(A,B) \times \mathcal{C}(B,C) &\xrightarrow{\circ} \mathcal{C}(A,C) \\ \mathcal{C}(A,B) \times \mathcal{C}(U,V) &\xrightarrow{\otimes} \mathcal{C}(A \otimes U, B \otimes V) \end{aligned}$$

Meaning of the sequential composition. The composite $A \xrightarrow{g \circ f} C$ inputs the cause $a \in A$ and outputs the effect $g(f(a)) \in C$ of the cause $f(a) \in B$, which is itself the effect of

the cause $a \in A$. In summary, we have

$$\begin{array}{c} a \in A \qquad f : A \longrightarrow B \\ f(a) \in B \qquad\qquad g : B \longrightarrow C \\ g \circ f(a) \;=\; g(f(a)) \in C \end{array} \tag{4}$$

Meaning of the parallel composition and product types. Since the strings in a string diagram correspond to types, drawing parallel strings leads to product types, like $A \otimes U$, which is the name of the type corresponding to the strings A and U running in parallel. The events of this type are the pairs $\langle a, u\rangle \in A \otimes U$, where $a \in A$ and $u \in U$. The parallel composite $A \otimes U \xrightarrow{f \otimes t} B \otimes V$ can thus be defined as the causal process transforming independent pairs of causes into independent pairs of effects, without any interferences between the components:

$$\begin{array}{cccc} \langle a, u\rangle \in A \otimes U & & \langle a, u\rangle \in A \otimes U & \\ a \in A \qquad f : A \longrightarrow B & & u \in U \qquad t : U \longrightarrow V & \\ f(a) \in B & & t(u) \in V & \\ (f \otimes t)\langle a, u\rangle = \big\langle f(a), t(u)\big\rangle \in B \otimes V & & & \end{array}$$

B Units

Vectors, scalars, covectors. There are processes where events occur with no visible causes; and there are processes where events have no visible effects. Such processes correspond, respectively, to string diagrams c and e in Fig. 11. There are even processes with no observable causes or effects, like the one represented by the diamond s in the middle of Fig. 11. When there are no strings at the bottom, or at the top of a box in a string diagram, we usually contract the bottom side, or the top side, into a point, and the box becomes a triangle. When there are no strings either at the bottom or at the top, then contracting both of them results in a diamond. A diamond with no inputs and outputs may still contain a lot of information. E.g., if causal processes are timed, then reading the time without observing anything else would be a process with no observable causes or effects. If processes are viewed as linear operators, then those that do not consume any input vectors and do not produce any output vectors are scalars. The processes that do not consume anything but produce output are just vectors, since they boil down to their output. The processes that do not produce vectors, but only consume them, are just covectors, or linear functionals.

Invisible string: The unit type I. Since every process must have an input type and an output type for formal reasons, in order to fit into a category of processes, a process that does not have any actual causes is presented in the form $I \xrightarrow{e} A$, where I is the *unit type*, satisfying

$$I \otimes A \;=\; A \;=\; A \otimes I \tag{5}$$

for every type A. The unit type is thus degenerate, in the sense that any number of its copies can be added to any type, without changing its elements. It is easy to see that it is unique, as the units in algebra tend to be. A process that does not output any effects is then in the form $A \xrightarrow{c} I$. The unit type is the unit with respect to the type product and to the parallel composition of processes, just like 0 is the unit with respect to the addition of numbers. It can be thought of as the type of a single event that never interferes with any other events.

Since it is introduced only to make sure that everything has a type, but otherwise has no visible causes or effects, the unit type is usually not drawn in string diagrams, but thought of as an *"invisible string"*. In Fig. 11, the invisible string is coming in below e, and out above c, and on both sides of s.

Figure 11: String diagrams with invisible strings

Invisible boxes: The unit processes id_A. For every type A there is a unit process $A \xrightarrow{\mathrm{id}_A} A$, called the *identity* of A, such that

$$\mathrm{id}_B \circ f \;=\; f \;=\; f \circ \mathrm{id}_A \tag{6}$$

holds for every process $A \xrightarrow{f} B$. This property is clearly analogous to (5), but has a different meaning: the causal process id_A inputs causes and outputs effects, both of type A; but it does not modify anything: it just outputs every cause $a \in A$ as its own effect. Since the box corresponding to id_A thus just passes to the output string whatever comes at the input string, and there is nothing else in it, we do not draw such a box. In string diagrams, the boxes corresponding to the identity processes id_A can thus be thought of as *"invisible boxes"*. In a sense, the strings themselves play the role of the identities.

Because of (5), any number of invisible strings can be added to any string diagram without changing its meaning. Because of (6), any number of invisible boxes can be added on any string, visible or invisible, without changing its meaning. This justifies eliding the units not only from string diagrams, but also from algebraic expressions, and writing things like

$$f \otimes U \;=\; f \otimes \mathrm{id}_U \qquad \text{and} \qquad A \otimes t \;=\; \mathrm{id}_A \otimes f$$

With this notation, the algebra of tensor products, studied in serious math, boils down to a single law, from which everything else follows:

C The middle-two-interchange law

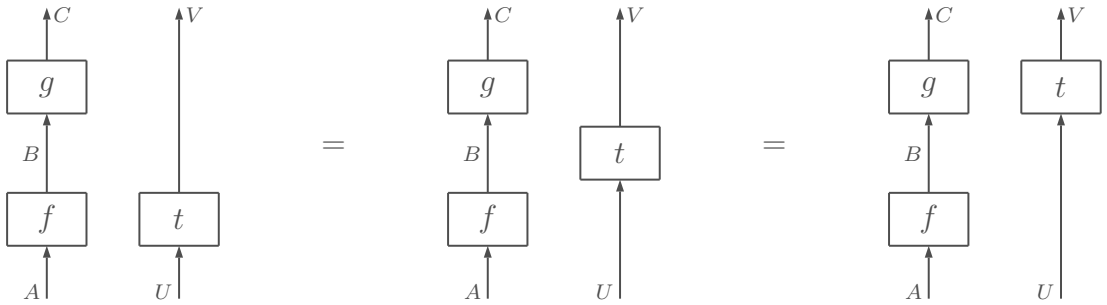
The main reason why string diagrams are convenient is that their geometry captures the *middle-two-interchange* law:

$$(f;g)\otimes(t;s) \quad = \quad (f\otimes t);(g\otimes s) \tag{7}$$

Note that the both sides of this equation correspond to the same string diagram, displayed in Fig. 10. The left-hand side of (7) is obtained by first reading the sequential compositions vertically, and then composing them in parallel horizontally; whereas the right-hand side is obtained by first reading the parallel compositions horizontally, and then composing them in sequence vertically. Sliding boxes along stings provides an easy way to derive further equations from the middle-two-interchange low of (7), such as

$$(f\otimes t);(g\otimes V) \qquad = \qquad (f\otimes U);(B\otimes t);(g\otimes V) \qquad = \qquad (f\otimes U);(g\otimes t)$$

In string diagrams, this is just



D Functions

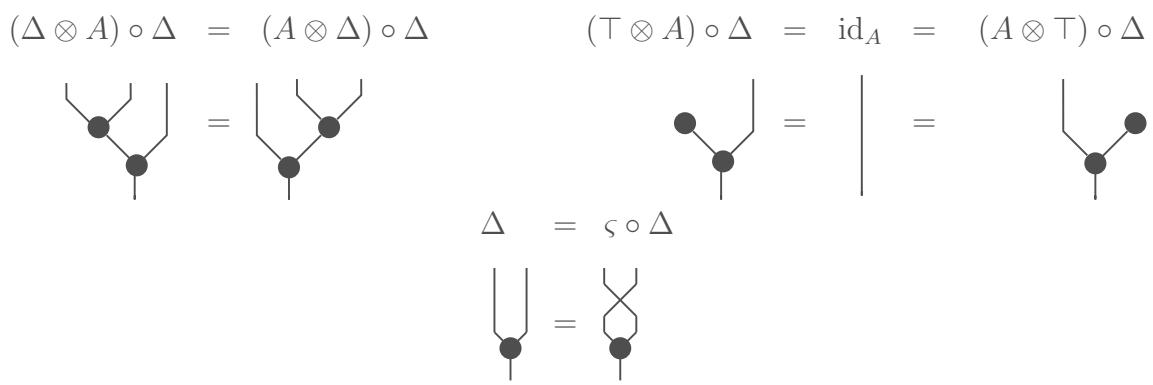
Causations are usually uncertain to some extent: a cause induces an effect with a certain probability. Those causations that are certain, so that the input always produces the output, and the same input always produces the same output, are called *functions*. If causations are presented as stochastic processes, then functions are the subfamily deterministic processes.

An intrinsic characterization of functions in a monoidal framewrok is provided by the following definitions.

A **data type** is a type $A\in\mathcal{C}$, equipped with *data services*, i.e. the operations

$$I \xleftarrow{\top} A \xrightarrow{\Delta} A\otimes A \tag{8}$$

respectively called *deleting* and *copying*, which satisfy the following equations:



We define *data* as the elements preserved by data services. Such elements are precisely those that can be manipulated using variables [30, 32].

Remark. If we dualize data services, i.e. reverse the arrows in (8), we get a binary operation and a constant. Transferring the equations from deleting and copying makes the binary operation associative and commutative, and it makes the constant into the unit. The dual of data services is thus the structure of a *commutative monoid*. The structure of data services itself is thus a commutative *co*monoid.

Functions are causations that map data to data. A causation $A \xrightarrow{f} B$ is a function if it is *total* and *single-valued*, which respectively corresponds to the two equations in Fig. 12. They also make f into a comonoid homomorphism from the data service comonoid on A to the data service comonoid on B.

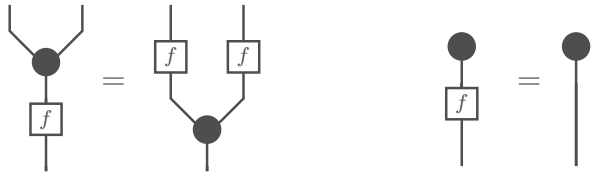


Figure 12: f is a function if and only if it is a comonoid homomorphism