Subjectivity, Bayesianism, and Causality

Pedro A. Ortega

***Abstract*—Bayesian probability theory is one of the elementary frameworks to model reasoning under uncertainty. Its defining property is the interpretation of probabilities as degrees of belief in propositions about the state of the world relative to an inquiring subject. This essay examines the notion of subjectivity in Bayesian probability theory. It turns out that the assumptions about subjectivity have a long tradition in Western culture that lie at the heart of its belief system, political organization and intellectual discourse. As an example, I show that some basic concepts of Bayesian probability theory have a counterpart in Lacanian theory. Furthermore, Lacanian theory explains agency in terms of an interruption of the signifying chain of the subject performed by the so called "objet petit a", which turns out to have striking similarities with causal interventions in statistical causality. Finally, an abstract model of subjective interaction is introduced that accommodates causal interventions in a measure-theoretic formalization.**

***Index Terms*—Subjectivity; Bayesian Probability Theory; Causality**

I. Introduction

EARLY modern thinkers of the Enlightenment, spurred by the developments of empirical science, modern political organization and the shift from collective religion to personal cults, found in the free, autonomous and rational *subject* the *locus* on which to ground all of knowledge (Foucault, 1966; Mansfield, 2000). Most notably, Descartes, with his axiom *cogito ergo sum* ('I think, therefore I am'), put forward the idea that the thought process of the subject is an unquestionable fact from which all other realities derive—in particular of oneself, and in general of everything else (Descartes, 1637).

This proposition initiated a long-lasting debate among philosophers such as Rousseau, Berkeley and Kant, and its discussion played a fundamental role in shaping modern Western civilisation. Indeed, the concept of the subject operates at the heart of our core institutions: the legal and political organization rests on the assumption of the free and autonomous subject for matters of responsibility of action and legitimization of ruling bodies; capitalism, the predominant economic system, depends on forming, through the tandem system of education and marketing, subjects that engage in work and consumerism (Burkitt, 2008); natural sciences equate objective truth with inter-subjective experience (Kim, 2005); and so forth.

Nowadays, questions about subjectivity are experiencing renewed interest from the scientific and technological communities. Recent technological advances, such as the availability of massive and ubiquitous computational capacity, the internet, and improved robotic systems, have triggered the proliferation of autonomous systems that monitor, process and deploy information at a scale and extension that is unprecedented in history. Today we have social networks that track user preferences and deliver personalized mass media, algorithmic trading systems that account for a large proportion of the trades at stock exchanges, unmanned vehicles that navigate and map unexplored terrain. What are the "users" that a social network aims to model? What does an autonomous system know, and what can it learn? Can an algorithm be held responsible for its actions? Furthermore, latest progress in neuroscience has both posed novel questions and revived old ones, ranging from investigating the neural bases of perception, learning and decision making, to understanding the nature of free will (Sejnowski and van Hemmen, 2006). Before these questions can be addressed in a way that is adequate for the mathematical disciplines, it is necessary to clarify what is meant by a subject in a way that enables a quantitative discussion.

The program of this essay is threefold. First, I will argue that Bayesian probability theory is a subjectivist theory, encoding many of our implicit cultural assumptions about subjectivity. To support this claim, I will show that some basic concepts in Bayesian probability theory have a counterpart in Lacanian theory, which is used in cultural studies as a conceptual framework to structure the discourse about subjectivity. In the second part, I will put forward the claim that Bayesian probability theory needs to be enriched with causal interventions to model agency. Finally, I will consolidate the ideas on subjectivity in an abstract mathematical synthesis. The main contribution of this formalization is the measure-theoretic generalization of causal interventions.

II. Subjectivity in Lacanian Theory

To artificial intelligence, statistics and economics, the questions about subjectivity are not novel at all: many can be traced back to the early discussions at the beginning of the twentieth century that eventually laid down the very foundations of these fields. Naturally, these ideas did not spring out of a vacuum, but followed the general trends and paradigms of the time. In particular, many of the fundamental concepts about subjectivity seem to have emerged from interdisciplinary cross-fertilization.

For instance, in the humanities, several theories of subjectivity were proposed. These can be roughly subdivided into two dominant approaches (Mansfield, 2000): the *subjectivist/psychoanalytic* theories, mainly associated with Freud and Lacan, which see the subject as a *thing* that can be conceptualized and studied (see e.g. Freud, 1899; Fink, 1996); and the *anti-subjectivist* theories, mainly associated with the works of Nietzsche and Foucault, which regard any attempt at defining the subject as a *tool* of social control, product of the culture and power of the time (Nietzsche, 1887; Foucault, 1964).

Email: ope@seas.upenn.edu. Address: School of Engineering and Applied Sciences, University of Pennsylvania, 220 S 33rd St, Philadelphia, PA 19104, U.S.A.

For our discussion, it is particularly useful to investigate the relation to Lacan[1], firstly because it is a subjectivist theory and secondly because its abstract nature facilitates establishing the relation to Bayesian probability theory. Some ideas that are especially relevant are the following.

The subject is a construct. There is a consensus among theorists (both subjectivist and anti-subjectivists) that the subject is not born into the world as a unified entity. Instead, its constitution as a unit is progressively built as it experiences the world (Mansfield, 2000). The specifics of this unity vary across the different accounts, but roughly speaking, they all take on the form of a separation of the sense of self (inside) from the rest of the world (outside). For instance, during the early stages of their lives, children have to learn that their limbs belong to them. In Lacan for instance, this distinction is embodied in the terms *I* and *the Other* (Fig. 1a). Crucially, Lacan stresses that the subject is precisely this "membrane" between inward and outward flow (Fig. 1a).

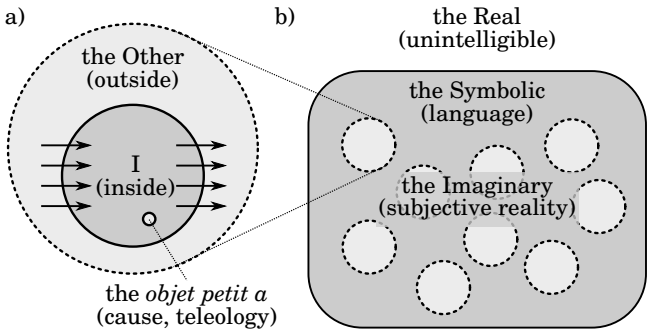


Fig. 1. The Subject in Lacanian Theory.

The subject is split. Structurally, the subject is divided into a part that holds beliefs about the world, and a part that governs the organization and dynamics of those beliefs *in an automatic fashion*. The most well-known instantiation of this idea is the Freudian distinction between the *conscious* and the *unconscious*, where the latter constitutes psychological material that is repressed, but nevertheless accessible through dreams and involuntary manifestations such as a "slip of the tongue" (Freud, 1899). Here however, the interpretation that is more interesting for our analysis is Lacan's. In his terminology, the two aforementioned parts correspond to the *imaginary* and the *symbolic* registers respectively (Fig. 1b). Simply put, the imaginary can be described as the collection of concepts or images that, when pieced together, make up the totality of the subject's ontology: in particular, the world and the subject's sense of self. In other words, the imaginary register is responsible for entertaining hypotheses about reality. In turn, these images are organized by the symbolic register into a network of meaning that is pre-given, static, and "structured like a language" (Lacan, 1977).

[1]It shall be noted however, that Lacan's work is notoriously difficult to understand, partly due to the complexity and constant revisions of his ideas, but most importantly due to his dense, multi-layered, and obscure prose style. As a result, the interpretation presented here is based on my own reading of it, which was significantly influenced by Fink (1996), Mansfield (2000) and the work by Žižek (Žižek, 1992, 2009).

Language is a system of signification. Many of the modern ideas about knowledge and subjectivity are centred around language. In this view, the subject is seen as a signifying entity that produces and consumes signs (linguistic material) in the form of spoken language, images, and general sensorimotor expression (de Saussure, 1916). Language then can be thought of as a system of signs that operates by detecting signifiers (labels) and associating them to signifieds (meanings or ideas)—possibly in cascade, with the signifieds being the signifiers of later stages. Crucially, the associations between signifiers and signifieds are arbitrary and contingent, established by pure convention (think of 'apple', 'manzana', 'mela', 'Apfel', 'pomme', 'תפוח', etc.). The influence of these views is witnessed by the adoption of related ideas by thinkers from fields ranging from logic (Russell, 1905; Wittgenstein, 1933), philosophy of language (Wittgenstein, 1953), phenomenology (Heidegger, 1927), rhetoric (Knape, 2000) and linguistics/cognitivism (Chomsky, 1957) to computer science (Turing, 1937) and biology/cybernetics (Maturana, 1970; Maturana and Varela, 1987).

The real is the engine of the subject. The imaginary and the symbolic registers refer to the subject's intellect, that is, to the organization of the things that it can potentially comprehend or experience. There is a third register in Lacan's conceptualization, namely the *real* (Fig. 1b), representing the unintelligible, random source of external perturbations that the subject picks up and integrates into its symbolic domain in the form of sense-data (compare e.g. to the "web of beliefs" of Quine (1951)).

Teleology. Finally, there is the question of purposeful behaviour. In Lacan, teleology is related to what he calls *objet petit a* (Fig. 1a), representing an interruption of the signifying chain inducing a lack of symbolic coherence (Lacan, 1973; Žižek, 1992). Such an interruption has two consequences that are worth pointing out. First, the deviation from the natural course of signification can be thought of as an expression of spontaneous desire marking a direction of behaviour in the sense of an instinct or drive (Freud, 1920). Second, the interrupted signifying chain, by injecting randomness, introduces an independence of choice that entails a responsibility, a claim to ownership of cause, and a post-rationalization of the subject's decisions. In short: a detected irregularity signals *agency*. For instance, in the sequence

$$1, \quad 2, \quad 3, \quad 4, \quad 5, \quad 6, \quad 8, \quad 9, \quad 10,$$

the missing number 7 breaks the pattern and can give the impression that it was intentionally omitted.

III. Subjectivity in Bayesian Probability Theory

In the mathematical disciplines, one of the most prominent theories dealing with subjectivity is Bayesian probability theory. Its current formal incarnation came to be as a synthesis of many fields such as measure theory (see e.g. Lebesgue, 1902; Kolmogorov, 1933), set theory (Cantor, 1874) and logic (Hegel, 1807; Frege, 1892; Russell, 1905; Wittgenstein, 1933). After Bayes' and Laplace's initial epistemic usage of probabilities (Bayes, 1763; Laplace, 1774), founders of modern

probability theory have *explicitly* started using probabilities as degrees of subjective belief. On one hand, they have postulated that subjective probabilities can be inferred by observing actions that reflect personal beliefs (Ramsey, 1931; De Finetti, 1937; Savage, 1954); on the other hand, they regarded probabilities as extensions to logic under epistemic limitations (Cox, 1961; Jaynes and Bretthorst, 2003). Importantly, both accounts rely on a subject that does statistics in the world having belief updates governed by Bayes' rule.

Bayesian probability theory, in its capacity as a subjectivist theory, can be related to ideas in Lacanian theory. Recall that formally, probability theory provides axiomatic foundations for modelling experiments involving randomness. Such a *randomized experiment* takes the form of a probability space $(\Omega, \mathcal{F}, \mathsf{P})$, where Ω is a set of possible states of nature, $\mathcal{F}$ is a σ-algebra on Ω (a collection of subsets of Ω that is closed under countably many set operations, comprising complement, union and intersection), and P is a probability measure over $\mathcal{F}$. Given this setup, I suggest the following correspondences, summarized in Tab. I:

1) *Real $\leftrightarrow$ generative/true distribution.* In probability theory, it is assumed that there exists a source that secretly picks the state of Nature $\omega \in \Omega$ that is then progressively "revealed" through measurements (Ash and Doléans-Dade, 1999). Some measure theory textbooks even allude to the irrational, unintelligible quality of the source[2] by using the phrase "Tyche, the goddess of chance, picks a sample" to describe this choice (see for instance Billingsley, 1978; Williams, 1991).
2) *Symbolic $\leftrightarrow$ probability space.* Conceptually, the σ-algebra $\mathcal{F}$ of a probability space contains the universe of all the yes/no questions (i.e. propositions) that the subject can entertain. A particular aspect of a given state of Nature ω is extracted via a corresponding random variable $X : \Omega \to \mathcal{X}$, mapping ω into a symbol $X(\omega)$ from a set of symbols $\mathcal{X}$. Random variables can be combined to form complex aspects, and the ensuing symbols are consistent (i.e. free of contradictions) as guaranteed by construction. Thus, a probability space and the associated collection of random variables make up the structure of the potential realities that the subject can hope to comprehend. Furthermore, one can associate to each random variable at least one of three roles (but typically just one), detailed next.
3) *Imaginary $\leftrightarrow$ hypotheses.* A random variable can play the role of a *latent feature* of the state of Nature. Latent variables furnish the sensorimotor space with a conceptual or signifying structure, and a particular configuration of these variables constitutes a hypothesis in the Bayesian sense. Because of this function, we can associate the collection of latent variables to Lacan's imaginary register.
4) *Flow between I and the Other $\leftrightarrow$ actions & observations.* The hypotheses by themselves do not ground the subject's symbolic domain to any reality however—for this, variables modelling interactions are required. These variables capture symbols that appear in the sensorimotor stream of the subject, that is, at its boundary to the world, modelling the directed symbolic flow occurring between the I and the Other; in particular, the out- and inward flows are represented by actions and observations respectively.
5) *Objet petit a $\leftrightarrow$ causal intervention.* The last connection I would like to establish, which will become a central theme in what follows, is between the *object petit a* and causal interventions. Lacanian theory explains agency in terms of a kink in the signifying chain—that is, the interruption of a pre-existing relation between two symbols—that is subjectivised *in hindsight* (Fink, 1996; Žižek, 1992). One crucial aspect of this notion is that it requires the comparison between two instants of the signifying network, namely the one where the relation is still intact and the resulting one where the relation is absent, adding a *dynamic* element to the static symbolic order. This element has *no* analogue in standard probability theory. However, the last twenty years have witnessed the systematic study of what appears to be an analogous idea in the context of probabilistic causality. More precisely, the interruption of the signifying chain is a causal intervention (Pearl, 2009).

TABLE I
LACANIAN AND BAYESIAN THEORIES OF THE SUBJECT

Lacan		Bayes
real (register)	$\longleftrightarrow$	true distribution
symbolic (register)	$\longleftrightarrow$	probability space
imaginary (register)	$\longleftrightarrow$	hypotheses
the Other $\to$ I (flow)	$\longleftrightarrow$	observations
I $\to$ the Other (flow)	$\longleftrightarrow$	actions
objet petit a	$\longleftrightarrow$	causal intervention

One can establish a few more connections, for instance between Lacan's concept of *jouissance* and the economic term *utility*, but I hope that the aforementioned ones suffice to make my case for now.

In summary, my claim is that Bayesian probability theory is almost an axiomatic subjectivist theory; "almost" because it lacks an analogue of the function performed by the *objet petit a*, namely causal interventions, which is crucial to fully characterize the subject. This will be the goal of the next section.

IV. PROBABILISTIC CAUSALITY

Causality has always been one of the central aspects of human explanation, with its first philosophical discussion dating back to Aristotle's *Physics* roughly some 2500 years ago. In spite of this, it has not received much attention from the scientific community, partly due to the strong scepticism expressed by Hume (1748) and later by prominent figures in statistics (Pearson et al., 1899) and logic (Russell, 1913). It is only in the recent decades that philosophers and computer

[2]Note that this allusion goes at least as far back as Hesiod's *Theogony* dating from the pre-philosophical era. The *Theogony* describes the creation of the world by the *muses*—a literary device used at the time standing for something that is unintelligible.

scientists have attempted to characterize causal knowledge in a rigorous way (Suppes, 1970; Salmon, 1980; Rubin, 1974; Cartwright, 1983; Spirtes and Scheines, 2001; Pearl, 2009; Woodward, 2013; Shafer, 1996; Dawid, 2007). I refer the reader to Dawid (2010) and references therein for a comparison between existing approaches.

Arguably, the central contribution has been Pearl's characterization of causal intervention (Pearl, 1993, 2009), which in turn draws ideas from Simon (1977) and Spirtes and Scheines (2001). Informally, a causal intervention is conceived as a manipulation of the probability law of a random experiment that functions by holding the value of a chosen random variable fixed. The operation has been formalized in causal directed acyclic graphs; structural equations (Pearl, 2009); chain graphs (Lauritzen and Richardson, 2002); influence diagrams (Dawid, 2007) and decision problems (Dawid, 2014), to mention some. Another approach that is worth pointing out is that of Shafer (1996). Therein, Shafer shows that simple probability trees are able to capture very rich causal structures, although he does not define causal interventions on them. While the aforementioned definitions differ in their scope, interpretation and simplicity, ultimately they entail transformations on probability measures that are mathematically consistent with each other.

Interestingly though, one the earliest and most general uses of what much later became known as causal interventions comes from game theory in the context of *extended games with imperfect information* (Osborne and Rubinstein, 1999). This connection is straightforward but, rather surprisingly, rarely acknowledged in the literature. The game-theoretic approach yields an elegant definition that is conceptually equivalent to defining Pearl-type interventions on Shafer's probability trees; furthermore, it lends itself to a measure-theoretic abstraction. For the sake of simplicity, this is the approach that I will adopt here.

A. Evidential versus Generative

One of the features of modern probability theory is that a probability space $(\Omega, \mathcal{F}, \mathsf{P})$ can be used in two ways, which we may label as *evidential* and *generative*. The evidential use conceives a probability space as a representation of an experimenter's knowledge about the conclusions he can draw when he is provided with measurements performed on the outcomes; while the generative usage sees the probability space as a faithful characterization of the experiment's stochastic mechanisms that bring about observable quantities. Thus, a statistician or a philosopher can use probabilities to asses the *plausibility* of hypotheses, while an engineer or a physicist typically uses them to characterize the *propensity* of an event, often assuming that these propensities are objective, physical properties thereof.

In a Bayesian interpretation, evidential and generative refer to subject's observations/measurements and actions/choices respectively. Under the evidential usage of probabilities, the subject passively contemplates the measurements of phenomena generated by the world. A measurement reveals to the subject which possible worlds he can discard from his knowledge state. In contrast, under the generative usage of probabilities, the subject *is* the random process itself. Thus, outcomes are chosen randomly by the subject and then communicated to the world. While there are many cases where this distinction does not play a role, if we aim at characterizing a subject that both passively observes and actively chooses, this distinction becomes crucial.

Our running example consists of a three-stage experiment involving two identical urns: the left one containing one white and three black balls, and the right one having three white and one black ball. In stage one, the two urns are either swapped or not with uniform probabilities. In stage two it is randomly decided whether to exclude the left or the right urn from the experiment. If the urns have not been swapped in the first stage, then the odds are $3/4$ and $1/4$ for keeping the left and the right urn respectively. If the urns have been swapped, then the odds are reversed. In the third and last stage, a ball is drawn from the urn with equal probabilities and its colour is revealed. We associate each stage with a binary random variable: namely Swap $\in \{\text{yes}, \text{no}\}$, Pick $\in \{\text{left}, \text{right}\}$ and Colour $\in \{\text{white}, \text{black}\}$ respectively. Figure 2 illustrates the setup. In calculations, I will sometimes abbreviate variable names and their values with their first letters. We will now consider several interaction protocols between two players named *I* and *W*, representing the outward and inward flow of a subject respectively as detailed in Section III.

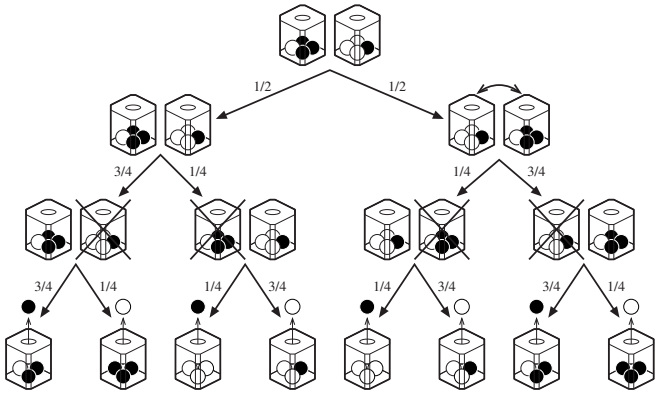


Fig. 2. A three-stage randomized experiment.

1) Generative: In the generative case, *I* carries out the three steps of the experiment, possibly consulting auxiliary randomizing devices like tosses of fair coins. In each step, *I* makes a random *choice*; that is, it selects a value for the corresponding random variable following a prescribed probability law that depends on the previous choices. For instance, the odds of "drawing a black ball" given that "the urns have been swapped in the first stage and the right urn has been picked in the second stage" is $3/4$.

The probabilities governing *I*'s behaviour are formalized with a probability space $S := (\Omega, \mathcal{F}, \mathsf{P})$, where Ω_1 contains the eight possible outcomes, σ-algebra is the powerset $\mathcal{F} = \mathscr{P}(\Omega_1)$, and P is the probability measure that is consistent with the conditional probabilities in Figure 2. Table II lists the eight outcomes and their probabilities.

The information contained in the probability space does not enforce a particular sequential plan to generate the outcome.

TABLE II
OUTCOME PROBABILITIES IN PROBABILITY SPACE S

Swap	Pick	Colour	Probability
no	left	black	9/32
no	left	white	3/32
no	right	black	1/32
no	right	white	3/32
yes	left	black	1/32
yes	left	white	3/32
yes	right	black	9/32
yes	right	white	3/32

The story of the experiment tells us that Swap, Pick and Colour are chosen in this order. However, *I* can construct other sequential plans to generate the outcome. For example, *I* could first choose the value of Colour, then Swap, and finally Pick (possibly having to change the underlying story about urns and balls), following probabilities that are in perfect accordance with the generative law specified by the probability space.

2) Evidential: In this case, player *I*, knowing about the probability law governing the experiment, passively observes its realisation as chosen by *W*. In each step, it makes a *measurement*; that is, *I* obtains the value of a random variable and uses it to update its beliefs about the state of the outcome. For instance, the plausibility of "the ball is black" given that "the urns have been swapped in the first stage and the right urn has been picked in the second stage" is $3/4$.

Here again, the probabilities governing *I*'s beliefs are formalized by the same probability space S. Analogously to the generative case, it does not matter in which order the information about the outcome is revealed to *I*: for instance, $\mathsf{P}(\text{Colour}|\text{Swap, Pick})$ is the same no matter whether it observes the value of Swap or Pick first.

B. Mixing Generative and Evidential

Let us change the experimental paradigm. Instead of letting *W* choosing and *I* passively observing, we now let both determine the outcome, taking turns in steering the course of the experiment depicted in Figure 2. In the first stage, *W* chooses between swapping the urns or not; in stage two, *I* decides randomly whether to keep the left or the right urn; and in the last stage, *W* draws a ball from the remaining urn. The protocol is summarized in Table III. We will investigate two experimental conditions.

TABLE III
PROTOCOL FOR THE EXPERIMENT

Stage	Variable	Chosen by
1	Swap	*W*
2	Pick	*I*
3	Colour	*W*

1) Perfect Information.: Under the first condition, both players are fully aware of all the previous choices. At any stage, the player-in-turn makes a decision following the conditional probability table that is consistent with past choices. It is easy to see that, again, the probability space S serves as a characterization of the subject: although this time the conditional probabilities stand for *I*'s beliefs (first and last stage) and *I*'s behaviour (second stage). Essentially, the fact that we now have interactions between *I* and *W* can still be dealt with under the familiar analytical framework of probability theory. Note that, as in the previous two cases, we can suggest changes to the sequential order; furthermore, we can swap the players' roles without changing our calculations.

2) Imperfect Information: The second experimental regime is identical to the previous one with one exception: *W* carries out the first stage of the experiment secretly, *without telling I* whether the urns were swapped or not. Hence, for *I*, the statements "Swap = yes" and "Swap = no" are equiprobable (e.g. the urns are opaque, see Fig. 3). How should *I* choose in this case? Let us explore two attempts.

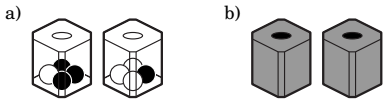


Fig. 3. Transparent versus opaque.

The first attempt consists in postulating that the two experimental regimes (perfect and imperfect information) are decoupled and hence require independent, case-based probability specifications. Concretely, $\mathsf{P}(\text{Pick}|\text{Swap} = \text{yes})$, $\mathsf{P}(\text{Pick}|\text{Swap} = \text{no})$ and $\mathsf{P}(\text{Pick})$ are independent probability distributions and are therefore freely specifiable. While this is a possible solution, it has the drawback that the resulting belief model violates the probability axioms, since

$$\mathsf{P}(\mathrm{P}) \neq \sum_{\mathrm{S}\,=\,\mathrm{y,\,n}} \mathsf{P}(\mathrm{P}|\mathrm{S})\,\mathsf{P}(\mathrm{S})$$

for almost all choices of the conditional probability distributions.

The second attempt enlarges the model as follows. We add an auxiliary binary variable, say KnowsSwap, that indicates whether *I* is in possession of the value of variable Swap. This allows specifying a total of four conditional probability distributions of the form

$$\mathsf{P}(\text{Pick}|\text{Swap}, \text{KnowsSwap})$$

indexed by the joint settings of Swap and KnowsSwap, where the latter can be treated as another random variable or as a conditional variable. However, this arrangement does not fundamentally bypass the original problem: we can extend *I*'s ignorance of the value of Swap to the value of KnowsSwap as well. That is, although we have extended Pick's functional dependency from Swap to both Swap and KnowsSwap, *there is no reason why KnowsSwap should not be an undisclosed variable too*. Consequently, this would require introducing yet another auxiliary variable, say KnowsKnowsSwap, to indicate whether the value of KnowsSwap is known, and so forth, piling up an infinite tower of indicator variables. Eventually, one is left with the feeling that this second solution is conceptually unsatisfactory as well.

Thus, let us continue with the game-theoretic solution to this problem, accepting *I*'s ignorance of the value of the

random variable Swap. The story of the experiment tells us that the probabilities $\mathsf{P}(\text{Pick}|\text{Swap})$ have the semantics of conditional instructions for *I*; but since the condition is unknown, the choice probabilities consistent with this situation are obtained by marginalizing over the unknown information. More specifically,

$$\mathsf{P}(\mathrm{P}=\mathrm{r}) = \sum_{\mathrm{S}\,=\,\mathrm{y,n}} \mathsf{P}(\mathrm{P}=\mathrm{r}|\mathrm{S})\,\mathsf{P}(\mathrm{S}) = \frac{3}{4}\cdot\frac{1}{2} + \frac{1}{4}\cdot\frac{1}{2} = \frac{1}{2}.$$

Interestingly, the resulting experiment does not follow the same generative law as in the previous experimental condition anymore, for the odds of swapping the urns in the first stage and picking the right urn in the second were $\frac{1}{2}\cdot\frac{3}{4} = \frac{3}{8}$ and not $\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{4}$ like in the current setup. Thus, albeit *I*'s beliefs are captured by the probability space $S = (\Omega, \mathcal{F}, \mathsf{P})$, the outcomes of this new experiment follow a different generative law, described by a probability triple $S' := (\Omega, \mathcal{F}, \mathsf{P}')$, where $\mathsf{P}' \neq \mathsf{P}$ is determined by the probabilities listed in Table IV. The choice made by *I* actually changed the probability law of the experiment!

TABLE IV
OUTCOME PROBABILITIES IN PROBABILITY SPACE S'.

Swap	Pick	Colour	Probability
no	left	black	3/16
no	left	white	1/16
no	right	black	1/16
no	right	white	3/16
yes	left	black	1/16
yes	left	white	3/16
yes	right	black	3/16
yes	right	white	1/16

A moment of thought reveals that this change happened simply because *I*'s state of knowledge did not conform to the functional requirements of the second random variable. At first seemingly harmless, this change of the probability law has far-reaching consequences for *I*'s state of knowledge: the familiar operation of probabilistic conditioning does not yield the correct belief update anymore. To give a concrete example, recall that the plausibility of the urns having been swapped in the first stage *before I* picks the left urn in the second stage is

$$\mathsf{P}(\mathrm{S}=\mathrm{y}) = \frac{1}{2}.$$

However, *after* the choice, the probability is

$$\mathsf{P}(\mathrm{S}=\mathrm{y}|\mathrm{P}=\mathrm{l}) = \frac{\mathsf{P}(\mathrm{P}=\mathrm{l}|\mathrm{S}=\mathrm{y})\mathsf{P}(\mathrm{S}=\mathrm{y})}{\sum_{\mathrm{S}\,=\,\mathrm{y,\,n}} \mathsf{P}(\mathrm{P}=\mathrm{l}|\mathrm{S})\mathsf{P}(\mathrm{S})} = \frac{\frac{1}{4}\cdot\frac{1}{2}}{\frac{1}{4}\cdot\frac{1}{2} + \frac{3}{4}\cdot\frac{1}{2}} = \frac{1}{4}$$

Hence, if *I* wanted to use probabilistic conditioning to infer the plausibility of the hypothesis, then it would conclude that its choice actually *created evidence* regarding the first stage of the experiment—a conclusion that violates common sense.

C. Causal Realisations

If we accept that probabilistic conditioning is not the correct belief update in the context of generative probabilities then we need to re-examine the nature of probabilistic choices.

The familiar way of conceptualizing the realisation of a random experiment $(\Omega, \mathcal{F}, \mathsf{P})$ is via the choice of a sample $\omega \in \Omega$ following the generative law specified by the probability measure P. Sequential observations are modelled as sequential refinements (i.e. a filtration)

$$\mathcal{F}_{\mathrm{I}} \xrightarrow{\mathrm{S}} \mathcal{F}_{\mathrm{II}} \xrightarrow{\mathrm{P}} \mathcal{F}_{\mathrm{III}} \xrightarrow{\mathrm{C}} \mathcal{F}$$

of an initial, ignorant algebra $\mathcal{F}_{\mathrm{I}} = \{\Omega, \varnothing\}$ up to the most fine-grained algebra $\mathcal{F} = \mathscr{P}(\Omega)$. The labels on the arrows indicate the particular random variable that has become observable (i.e. measurable) in the refinement. A second, non-standard way in the context of probability theory, is to think of a realisation as a *random transformation* of an initial probability triple $(\Omega, \mathcal{F}, \mathsf{P})$ into a final, degenerate probability triple $(\Omega, \mathcal{F}, \mathsf{P}_\omega)$, where P_ω is the probability measure concentrating all its probability mass on the singleton $\{\omega\} \in \mathcal{F}$. This alternative way of accounting for random realisations will prove particularly fruitful to formalize probabilistic choices.

In many situations it is natural to subdivide the realisation of a complex experiment into a sequence of realisations of simple sub-experiments. For instance, the realisation of the experiment in the running example can be broken down into a succession of three random choices, i.e. a sequence

$$\mathsf{P} \xrightarrow{f_{\mathrm{S}}} \mathsf{P}_{\mathrm{I}} \xrightarrow{f_{\mathrm{P}}} \mathsf{P}_{\mathrm{II}} \xrightarrow{f_{\mathrm{C}}} \mathsf{P}_{\mathrm{III}},$$

of random transformations of the initial probability measure P. Here, the three mappings f_{S}, f_{P} and f_{C} implement particular assignments for the values of Swap, Pick and Colour respectively, and P_{I}, P_{II} and $\mathsf{P}_{\mathrm{III}}$ are their corresponding resulting probability measures. Together, f_{S}, f_{P} and f_{C} form a *sequential plan* to specify a particular realisation $\{\omega\} \in \mathcal{F}$ of the experiment. However, the mathematical formalisation of such decompositions requires further analysis.

Underlying any purely evidential usage of probabilities, there is the implicit, although somewhat concealed, assumption of a predetermined outcome: the choice of the outcome of the experiment *precedes* the measurements performed on it (Here we recall the probabilistic mythology, in which it is *Tyche, the Goddess of Chance*, who has the privilege of choosing the outcome). In other words, obtaining information about the outcome updates the belief state of the subject, but not the outcome itself. In contrast, the generative use assumes an undetermined, fluid state of the outcome. More specifically, a choice updates both the beliefs of the subject *and the very state of the realisation*. Hence, there are two types of states that have to be distinguished: the state of the beliefs and the state of the realisation.

Distinguishing between the states of the realisation imposes restrictions on how beliefs have to be updated after making choices. These restrictions are probably best highlighted if one imagines—for illustrative purposes—that the experiment is a physical system made up from a cascade of (stochastic) mechanisms, where each mechanism is a sub-experiment implementing a choice. Based on the physical metaphor, one concludes that: (i) every choice can have past and future choices; (ii) a choice can depend on past, but not on future choices; (iii) a choice cannot change past choices; and (iv)

a choice can influence future choices. Or, put concisely: *the range of the effect of a choice is subject to its causal constraints*. So, for instance, picking the left urn knowing that the urns were swapped in the first stage increases the odds for drawing a white ball in the last stage, but it cannot change the fact that the urns have been swapped. Hence, the belief update following a choice affects the beliefs about the future, but not about the past[3].

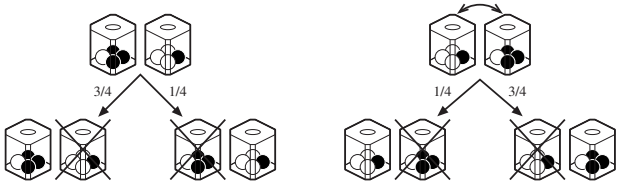


Fig. 4. The second stage of the randomized experiment.

Another aspect of choices is concerned with their scope, which spans many potential realisations. Consider the second stage of the experiment. As it can be seen from its illustration in Figure 4, this stage contains two parts, namely one for each possible outcome of the first stage. In a sense, its two constituents could be conceived as being actually two stand-alone experiments deserving to be treated separately, since they represent alternative, mutually exclusive historical evolutions of the sequential experiment which were rendered causally independent by the choice in the first stage. Thus, picking an urn given that the urns were swapped in the first stage could in principle have nothing to do with picking an urn given that the urns were not swapped. However, this separation requires knowing the choices in the execution of the sequential experiment; a situation that failed to happen in our previous example. To be more precise: the semantics of this grouping is precisely that we have declared not being able to discern between them. In game-theory, this is called an *information set*.

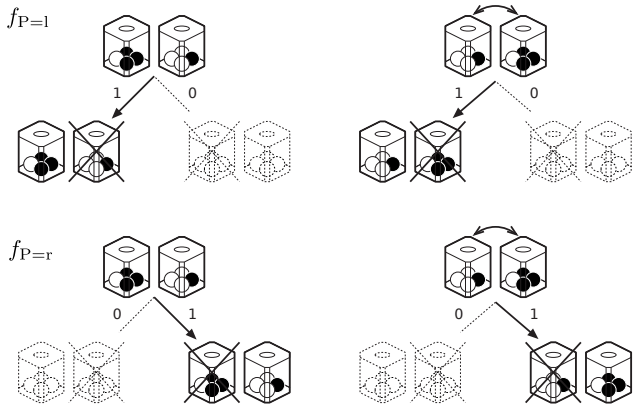


Fig. 5. The two possible choices in the second stage.

Even though it is clear that the belief update for a choice has to respect the causal boundaries separating the different histories, a choice is an epistemic operation that affects *all* histories in parallel because they are the *same* from the subject's point of view. Therefore, we formalize the sub-experiment in Figure 4 as a *collection* of experiments admitting two choices, namely $f_{\mathrm{P=l}}$ and $f_{\mathrm{P=r}}$, representing the choice of the left and right urn respectively. Suppose we are asked to choose the left urn in this collection of experiments. This makes sense because "choosing the left urn" is an operation that is well-defined across all the members in the collection. Then, this choice amounts to a transformation that puts all the probability mass on both left urns. Analogously, "choosing the right urn" puts all the probability mass on the right urns. The two choices, $f_{\mathrm{P=l}}$ and $f_{\mathrm{P=r}}$, are illustrated in Figure 5.

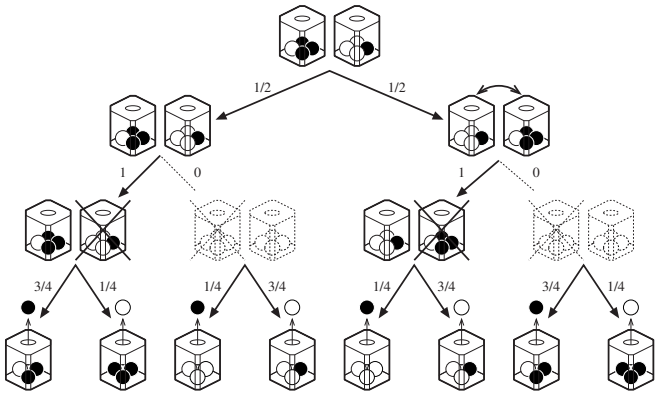


Fig. 6. The three-stage randomized experiment after choosing the left urn in the second stage.

Recall the situation where player *I* chose the left urn without knowing whether the urns were swapped or not in the first stage. From our previous discussion, this amounts to applying $f_{\mathrm{P=l}}$ to the sub-experiment in the second stage. This leads to the modified three-stage experiment illustrated in Figure 6 having a probability measure $\mathsf{P}_{\mathrm{P=l}}$, where the subscript informs us of the manipulation performed on the original measure P. Similarly, if $f_{\mathrm{P=r}}$ is applied, we obtain the probability measure $\mathsf{P}_{\mathrm{P=r}}$ for the experiment. Table V lists both probability measures plus the expected probability measure $\mathsf{E}[\mathsf{P}_{\mathrm{P}}]$ which averages over the two choices. Notice that in Table IV, it is seen that $\mathsf{E}[\mathsf{P}_{\mathrm{P}}]$ is equal to P', i.e. the probability law resulting from the experiment under the condition of imperfect information.

TABLE V
PROBABILITIES OF THE EXPERIMENT AFTER THE CHOICE OF PICK.

Swap	Pick	Colour	$\mathsf{P}_{\mathrm{P=L}}$	$\mathsf{P}_{\mathrm{P=R}}$	$\mathsf{E}[\mathsf{P}_{\mathrm{P}}]$
no	left	black	3/8	0	3/16
no	left	white	1/8	0	1/16
no	right	black	0	1/8	1/16
no	right	white	0	3/8	3/16
yes	left	black	1/8	0	1/16
yes	left	white	3/8	0	3/16
yes	right	black	0	3/8	3/16
yes	right	white	0	1/8	1/16

[3]To be more precise, by the terms *past* and *future* I mean the causal precedents and causal successors respectively.

Finally, we calculate the plausibility of the urns having been swapped after the left urn is chosen. This is given by

$$\begin{aligned}\mathsf{P}_{\mathrm{P}=\mathrm{l}}(\mathrm{S}=\mathrm{y}|\mathrm{P}=\mathrm{l}) &= \frac{\mathsf{P}_{\mathrm{P}=\mathrm{l}}(\mathrm{P}=\mathrm{l}|\mathrm{S}=\mathrm{y})\mathsf{P}_{\mathrm{P}=\mathrm{l}}(\mathrm{S}=\mathrm{y})}{\sum_{\mathrm{s}=\mathrm{y,n}}\mathsf{P}_{\mathrm{P}=\mathrm{l}}(\mathrm{P}=\mathrm{l}|\mathrm{S}=\mathrm{s})\mathsf{P}_{\mathrm{P}=\mathrm{l}}(\mathrm{S}=\mathrm{s})}\\ &= \frac{1\cdot\mathsf{P}(\mathrm{S}=\mathrm{y})}{\sum_{\mathrm{s}=\mathrm{y,n}}1\cdot\mathsf{P}(\mathrm{S}=\mathrm{s})}\\ &= \frac{1\cdot\frac{1}{2}}{1\cdot\frac{1}{2}+1\cdot\frac{1}{2}} = \frac{1}{2}.\end{aligned}$$

Hence, according to the belief update for choices that we have proposed in this section, choosing the left urn in the second stage does not provide evidence about the first stage. However, if a black ball is drawn right afterwards, the posterior plausibility will be

$$\begin{aligned}&\mathsf{P}_{\mathrm{P}=\mathrm{l}}(\mathrm{S}=\mathrm{y}|\mathrm{P}=\mathrm{l},\mathrm{C}=\mathrm{b})\\ &\quad= \frac{\mathsf{P}_{\mathrm{P}=\mathrm{l}}(\mathrm{C}=\mathrm{b}|\mathrm{S}=\mathrm{y},\mathrm{P}=\mathrm{l})\mathsf{P}_{\mathrm{P}=\mathrm{l}}(\mathrm{S}=\mathrm{y}|\mathrm{P}=\mathrm{l})}{\sum_{\mathrm{s}=\mathrm{y,n}}\mathsf{P}_{\mathrm{P}=\mathrm{l}}(\mathrm{C}=\mathrm{b}|\mathrm{S}=\mathrm{s},\mathrm{P}=\mathrm{l})\mathsf{P}_{\mathrm{P}=\mathrm{l}}(\mathrm{S}=\mathrm{s}|\mathrm{P}=\mathrm{l})}\\ &\quad= \frac{\frac{1}{4}\cdot\frac{1}{2}}{\frac{1}{4}\cdot\frac{1}{2}+\frac{3}{4}\cdot\frac{1}{2}} = \frac{1}{4},\end{aligned}$$

i.e. the subject obtains evidence favouring the hypothesis that the urns were not swapped. This leads to an interesting interpretation. In a sense, the intervention functions as a psychological mechanism informing the subject that its choice cannot be used as additional evidence to support hypotheses about the past; the fundamental role of the intervention is to declare the choice as an unequivocal, deterministic consequence of the subject's state of knowledge at the moment of the decision. Or, loosely speaking, the intervention "tricks the subject into believing that its choice was deliberate, not originating from an external source"—recall the discussion about the *objet petit a*. As a consequence, a subject can never learn from its own actions; rather, it only learns from their effects.

V. The Abstract Subject

Perhaps the two most important insights of our previous discussion are that, firstly, an interactive subject must distinguish between its *actions* and *observations* because they entail different belief updates, and secondly, that in order to do so, the subject must differentiate between an *event* (i.e. a logical proposition about the experiment) and a *realisation* (i.e. a possible state of the experiment). We have seen that the subject's actions directly shape the course of the realisation of the experiment. The subject, in order to understand and eventually predict the effects of an action, must intervene his or her beliefs about the state of the realisation. This intervention cannot be performed unless the subject is in possession of a description of the random experiment that enumerates its states of realisation and details their causal dependencies.

The aim of this section is to present an abstract model of the subject. In particular, I have dedicated much effort into elucidating the links to measure-theoretic probability, which currently holds the status of providing the standard foundations for abstract probability theory.

A. Realisations and Causal Spaces

First we introduce a structure that models the states of realisation of a random experiment.

Definition 1 (Realisation)**.** A set $\mathcal{R}$ of non-empty subsets of Ω is called a *set of realisations* iff

- R1. *the sure event is a realisation:* $\Omega \in \mathcal{R}$;
- R2. *realisations form a tree:* for each distinct $U, V \in \mathcal{R}$, either $U \cap V = \varnothing$ or $U \subset V$ or $V \subset U$;
- R3. *the tree is complete:* for each $U, V \in \mathcal{R}$ where $V \subset U$, there exists a sequence $(V_n)_{n\in\mathbb{N}}$ in $\mathcal{R}$ such that $U \setminus V = \bigcup_n V_n$.
- R4. *every branch has a starting and an end point:* let $(V_n)_{n\in\mathbb{N}} \in \mathcal{R}$ be such that $V_n \uparrow V$ or $V_n \downarrow V$. Then, $V \in \mathcal{R}$.

A member $U \in \mathcal{R}$ is called a *realisation* or a *realisable event*. Given two realisations $U, V \in \mathcal{R}$, we say that U *precedes* V iff $U \supset V$. Given two subsets $\mathcal{U}, \mathcal{V} \subset \mathcal{R}$, we say that $\mathcal{U}$ *precedes* $\mathcal{V}$ iff for every $V \in \mathcal{V}$, there exists an element $U \in \mathcal{U}$ such that U precedes V. Analogously, we also say that V *follows* U iff U precedes V, and that $\mathcal{V}$ *follows* $\mathcal{U}$ iff $\mathcal{U}$ precedes $\mathcal{V}$. Finally, two realisations $U, V \in \mathcal{R}$ that neither precede nor follow each other are said to be *incomparable*.

From axioms R1–R3, it is clearly seen that a set of realisations is essentially a tree of nested subsets of the sample space, rooted at the sample space. Axiom R4 includes the upper and lower limits of realisation sequences, thus constituting a formalisation of the fourth postulate causal reasoning. One important difference to standard σ-algebras is that the complement is, in general, *not* in the algebra, the only exception being the impossible realisation.

An immediate consequence of this definition is that the set of realisations forms a partial order among its members. The partial order is the fundamental requirement for modelling causal dependencies.

Proposition 2 (Partial Order)**.** *A set of realisations $\mathcal{R}$ endowed with the set inclusion $\subset$ forms a partial order.*

Proof. Trivial, because it is inherited from $(\mathscr{P}(\Omega), \subset)$: it is reflexive, since for each $U \in \mathcal{R}$, $U \subset U$; it is antisymmetric, since for each $U, V \in \mathcal{R}$, if $U \subset V$ and $V \subset U$ then $U = V$; and it is transitive, because for all $U, V, W \in \mathcal{R}$, if $W \subset V$ and $V \subset U$ then $W \subset U$. □

The intuition here is that "$U \supset V$" corresponds to the intuitive notion of "V depends causally on U", i.e. the veracity of V can only be determined insofar U is known to have obtained; and "$U \cap V = \varnothing$" means that "V and U are causally independent".

A set of realisations can be visualised as a tree of nested sets. For instance, Fig. 7 is a possible set of realisations for the experiment in Fig. 2. Here, the sure event Ω at the root is partitioned recursively into branches until reaching the leaves representing the termination of the experiment.

Next we define an important class of events of the experiment, namely those that can be thought of as a union of causal

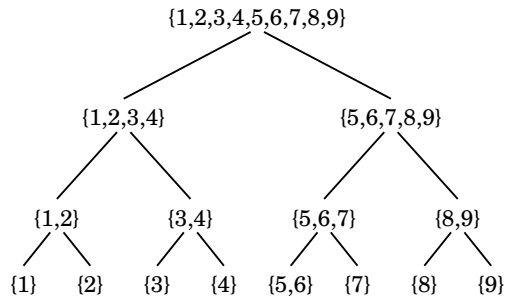


Fig. 7. A realisation set for the experiment.

histories that are potentially incompatible.

Definition 3 (Representation)**.** A subset $A \subset \Omega$ is said to have a *representation* in $\mathcal{R}$ iff there exists a sequence $(A_n)_{n\in\mathbb{N}}$ in $\mathcal{R}$ such that $\bigcup_n A_n = A$. The *set of representable events* $r(\mathcal{R})$ is the collection of all the subsets of Ω that have a representation in $\mathcal{R}$.

For instance, consider the subsets

$$A_1 = \{2, 5, 6\}, \quad A_2 = \{3, 4\} \quad \text{and} \quad A_3 = \{4, 5\}.$$

A_1 has a unique representation given by $\mathcal{A}_1 = \{\{2\}, \{5, 6\}\}$; A_2 has two representations, namely $\mathcal{A}_2 = \{\{3\}, \{4\}\}$ and $\mathcal{A}_2' = \{\{3, 4\}\}$; and A_3 does not have a representation.

It turns out that the set of representable events has a fundamental property: it coincides with the σ-algebra generated by $\mathcal{R}$. This means that every event of the experiment can be thought of as corresponding to a collection of possibly mutually exclusive realisations.

Theorem 4 (Representation)**.** *Let $\mathcal{R}$ be a set of realisations, let $r(\mathcal{R})$ be the set of representable events in $\mathcal{R}$ and let $\sigma(\mathcal{R})$ be the σ-algebra generated by $\mathcal{R}$. Then, $r(\mathcal{R}) = \sigma(\mathcal{R})$.*

Proof. Case $r(\mathcal{R}) \subset \sigma(\mathcal{R})$: This follows directly from the definition of a σ-algebra. *Case $r(\mathcal{R}) \supset \sigma(\mathcal{R})$:* We prove this by induction. For the base case, let $(V_n)_{n\mathbb{N}}$ be a sequence in $\mathcal{R}$. Then, $V = \bigcup_n V_n \in \sigma(\mathcal{R})$ has a representation in $\mathcal{R}$. Furthermore, $V \in \mathcal{R}$ implies that there exists $(V_n)_{n\in\mathbb{N}}$ in $\mathcal{R}$ such that $\Omega \setminus V = \bigcup_n V_n$ (Axiom R3). Hence, $V^c \in \sigma(\mathcal{R})$ too has a representation in $\mathcal{R}$. For the induction case, assume we have a sequence $(A_n)_{n\in\mathbb{N}}$ with representations $(\mathcal{A}_n)_{n\in\mathbb{N}}$ respectively, where $\mathcal{A}_n = (V_{n,m})_{m\in\mathbb{N}}$ for each $n \in \mathbb{N}$. Then,

$$\bigcup_n A_n = \bigcup_n \bigcup_m V_{n,m} = \bigcup_l V_l,$$

where $l \in \mathbb{N}$ is a diagonal enumeration of the $(n, m) \in \mathbb{N}\times\mathbb{N}$. Obviously, $(V_l)_{l\in\mathbb{N}}$ is a representation for $\bigcup_n A_n \in \sigma(\mathcal{R})$. Now, assume that $A \in \sigma(\mathcal{A})$ has a representation $(A_n)_{n\in\mathbb{N}}$. Then,

$$A^c = \left(\bigcup_n A_n\right)^c = \bigcap_n A_n^c.$$

Since the A_n are in $\mathcal{R}$, their complements A_n^c have representations $(V_{n,m})_{m\in\mathbb{N}}$. Hence,

$$A^c = \bigcap_n A_n^c = \bigcap_n \bigcup_m V_{n,m} = \bigcup_{f:\mathbb{N}\to\mathbb{N}} \bigcap_n V_{n,f(n)},$$

where the last equality holds due to the extensionality property of sets. More specifically, for $\omega \in \Omega$ to be a member of the l.h.s., there must be an m for each n such that $\omega \in V_{n,m}$. This is true in particular for the map f that chooses the smallest m for each n. Hence, ω is a member of the r.h.s. Now, consider an element $\omega \in \Omega$ that is not in the l.h.s. Then, there exists some n such that $\omega \notin V_{n,m}$ for all m. Since, for this particular n, this is false for any choice of f in the r.h.s., ω is not a member of the r.h.s., which proves the equality. Finally, since intersections of members $V_{n,m}$ of $\mathcal{R}$ are either equal to $\varnothing$ or equal to a member V_l of $\mathcal{R}$ (Axiom R2), one has $A^c = \bigcup_l V_l$ for some $(V_l)_{l\in\mathbb{N}}$, which is a representation of A^c. □

Having defined the basic structure of realisations, we now place probabilities on them. However, rather than working with the standard (unconditional) probability measure P that is customary in measure-theory, here—as is also the case in Bayesian probability theory (Cox, 1961; Jaynes and Bretthorst, 2003)—it is much more natural to work directly with a conditional probability measure $\mathsf{P}(\cdot|\cdot)$. To establish the connection to the standard measure-theoretic view, one can simply think of the conditional probability measure as defined by: $\mathsf{P}(A|\Omega) := \mathsf{P}(A)$; and in general $\mathsf{P}(A|U)$, for $A \in \sigma(\mathcal{R})$, $U \in \mathcal{R}$, as a version of the conditional expectation function for the indicator function $\mathsf{P}(A|U) := \mathsf{E}(1_A|\mathcal{G})$, where $\mathcal{G}$ is the algebra generated by U and all its predecessors. Henceforth, we will drop the qualifier "conditional", and just talk about the "probability measure" $P(\cdot|\cdot)$.

Definition 5 (Causal Measure)**.** Given a set of realisations $\mathcal{R}$, a *causal measure* is a binary set function $\mathsf{P}(\cdot|\cdot) : \mathcal{R}\times\mathcal{R} \to [0, 1]$, such that

C1. *the past is certain:*
For any $V, U \in \mathcal{R}$, if U precedes V, then

$$\mathsf{P}(U|V) = 1;$$

C2. *incomparable realisations are impossible:*
For any incomparable $V, U \in \mathcal{R}$,

$$\mathsf{P}(V|U) = 0;$$

C3. *sum-rule:*
For any $U \in \mathcal{R}$ and any disjoint sequence $(V_n)_{n\in\mathbb{N}}$ such that V_n follows U for all $n \in \mathbb{N}$ and $\bigcup_{n\in\mathbb{N}} V_n = U$,

$$\sum_{n\in\mathbb{N}} \mathsf{P}(V_n|U) = 1;$$

C4. *product-rule:*
For any $U, V, W \in \mathcal{R}$ such that W follows V and V follows U,

$$\mathsf{P}(W|U) = \mathsf{P}(W|V) \cdot \mathsf{P}(V|U).$$

Thus, a causal measure is defined only over $\mathcal{R}\times\mathcal{R}$, providing a supporting skeleton for the construction of a full-fletched probability measure extending over the entire σ-algebra. A simple way of visually representing a causal measure is by indicating the transition probabilities in the corresponding tree diagram, as illustrated in Fig. 8. In the figure, the sets have been replaced with labels, e.g. $S_0 = \Omega$ and $S_4 = \{3, 4\}$.

Definition 6 (Compatible Probability Measure)**.** Given a causal measure P over a set of realisations $\mathcal{R}$, a probability measure $\mathsf{P}'(\cdot|\cdot) : \sigma(\mathcal{R}) \times \sigma(\mathcal{R}) \to [0,1]$ is said to be *compatible* with P iff $\mathsf{P}' = \mathsf{P}$ on $\mathcal{R} \times \mathcal{R}$.

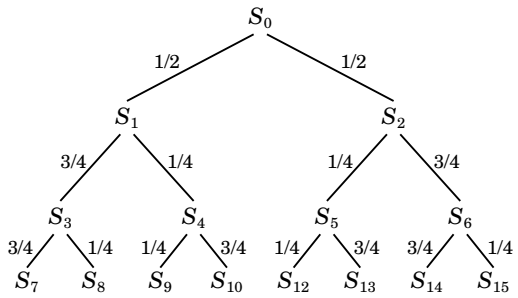


Fig. 8. Visualisation A causal space for the experiment.

It turns out that the causal measure almost completely determines its compatible probability measures, the exception being the probabilities conditioned on events that are not realisations. To show this, we first introduce a definition.

Definition 7 ($\mathcal{R}_U$, Σ_U)**.** Let $\mathcal{R}$ be a set of realisations. For any given U, define $\mathcal{R}_U := U \cap \mathcal{R}$ and $\Sigma_U := U \cap \sigma(\mathcal{R})$.

Observe that $\mathcal{R}_U$ is a set of realisations based on U as the sample space. Furthermore, it is well-known (Ash and Doléans-Dade, 1999, Chapter 1.2) that

$$\Sigma_U = U \cap \sigma(\mathcal{R}) = \sigma_U(U \cap \mathcal{R})$$

where $\sigma_U(U \cap \mathcal{R})$ is the σ-algebra generated by subsets of U, i.e. where U rather than Ω is taken as the sample space. The aforementioned uniqueness result follows.

Proposition 8. *Let P_1 and P_2 be two probability measures that are compatible with a causal measure P over $\mathcal{R}$. Then, for each $U \in \mathcal{R}$, $V \in \Sigma_U$, $\mathsf{P}_1(V|U) = \mathsf{P}_2(V|U)$.*

Proof. First we note that each $\mathcal{R}_U$ is a π-system, i.e. a family of subsets of U that is stable under finite intersection: $U, V \in \mathcal{R}_U$ implies $U \cap V \in \mathcal{R}_U$. This is because, for any $U \in \mathcal{R}_U$, $U \cap U \in \mathcal{R}_U$; and for all distinct $U, V \in \mathcal{R}_U$, either $U \cap V = \varnothing$ or $U \subset V$ or $V \subset U$ implies that either $U \cap V = \varnothing = U \setminus U$ or $U \cap V = U$ or $V \cap U = V$, which are all members of $\mathcal{R}_U$.

Next we prove that for each $U \in \mathcal{R}$, $V \in \Sigma_U$, $\mathsf{P}_1(V|U) = \mathsf{P}_2(V|U)$. Lemma 1.6. in Williams (1991) states that, if two probability measures agree on a π-system, then they also agree on the σ-algebra generated by the π-system. Pick any $U \in \mathcal{R}$. Applying the lemma, we conclude that for all $V \in \Sigma_U$, $\mathsf{P}_1(V|U) = \mathsf{P}_2(V|U)$. Since $U \in \mathcal{R}$ is arbitrary, the statement of the proposition is proven. □

Given the previous definition, we are ready to define our main object: the causal space. A causal space, like a standard probability space, serves the purpose of characterizing a random experiment, but with the important difference that it also contains information about the causal dependencies among the events of the experiment.

Definition 9 (Causal Space)**.** A *causal space* is a tuple $C = (\Omega, \mathcal{R}, \mathsf{P})$, where Ω is a set of outcomes, $\mathcal{R}$ is a set of realisations on Ω, and P is a causal measure over $\mathcal{R}$.

Intuitively, it is clear that a causal space contains enough information to characterize probability spaces that represent the same experiment. These probability spaces are defined as follows.

Definition 10 (Compatible Probability Space)**.** Given a causal space $C = (\Omega, \mathcal{R}, \mathsf{P})$, a probability space $S = (\Omega, \mathcal{F}, \mathsf{P}')$ is said to be compatible with C if $\mathcal{F} = \sigma(\mathcal{R})$ and P' is *compatible* with P.

An immediate consequence of the previous results is that compatible probability spaces are essentially unique.

Corollary 11. *Let $S_1 = (\Omega, \mathcal{F}_1, \mathsf{P}_1)$ and $S_2 = (\Omega, \mathcal{F}_2, \mathsf{P}_2)$ be two probability spaces compatible with a given causal space $C = (\Omega, \mathcal{R}, \mathsf{P})$. Then,*

1) *their σ-algebras are equal, i.e. $\mathcal{F}_1 = \mathcal{F}_2$;*
2) *and their probability measures are equal on any condition $U \in \mathcal{R}$ and σ-algebras Σ_U, i.e. for any $U \in \mathcal{R}$, $V \in \Sigma_U$, $\mathsf{P}_1(V|U) = \mathsf{P}_2(V|U)$.*

Importantly though, one cannot derive a unique causal space from a probability space; that is to say, given a probability space, there is in general more than one causal space that can give rise to it. Crucially, these causal spaces can differ in the causal dependencies they enforce on the events of the experiment, thus representing incompatible causal realisations.

B. Causal Interventions

The additional causal information contained in causal spaces serve the purpose of characterizing cause-effect relations (e.g. functional dependencies) and the effects of interventions of a random experiment. Interventions modify realisation processes in order to steer the outcome of a random experiment into a desired direction.

We begin this subsection with the formalisation of a subprocess as a sequence of realisations of the experiment, that is, as a realisation *interval*.

Definition 12 (Interval)**.** Let $U, V \in \mathcal{R}$. Define

$$\mathcal{I} := \{W \in \mathcal{R} : U \supset C \text{ and } C \supset V\}.$$

Then, based on $\mathcal{I}$, define:

$$\begin{aligned}
[U,V]_{\mathcal{R}} &:= \mathcal{I}, && \text{(closed interval)}\\
(U,V]_{\mathcal{R}} &:= \mathcal{I} \setminus \{U\}, && \text{(right-closed interval)}\\
[U,V)_{\mathcal{R}} &:= \mathcal{I} \setminus \{V\}, && \text{(left-closed interval)}\\
(U,V)_{\mathcal{R}} &:= \mathcal{I} \setminus \{U,V\}. && \text{(open interval)}
\end{aligned}$$

Obviously, for all $U, V \in \mathcal{R}$,

$$[U,V]_{\mathcal{R}} \neq \varnothing \qquad \Leftrightarrow \qquad U \supset V,$$

so that non-empty intervals necessarily have a causal direction. Although the previous definition covers open, half-open and closed intervals, we will see further down that only closed intervals play an important rôle in the context of interventions.

For example, in Fig. 8 we have:

$$[S_0, S_{12}]_{\mathcal{R}} = \{S_0, S_2, S_5, S_{12}\}$$
$$[S_0, S_{12})_{\mathcal{R}} = \{S_0, S_2, S_5\}$$
$$[S_{12}, S_0]_{\mathcal{R}} = \varnothing.$$

In a random experiment, two sub-processes with the same initial conditions can lead to two different outcomes. Next, I define a *bifurcation* and a *discriminant*, the former corresponding to the exact moment when these two processes separate from each other and the latter to the instant right afterwards—that is, the instant that unambiguously determines the start of a new causal course. Notice that in what follows, I will drop the subscript $\mathcal{R}$ when it is clear from the context.

Definition 13 (Bifurcations & Discriminants)**.** Let $\mathcal{R}$ be a set of realisations, and let $\mathcal{I}_1 = [U, V_1]$ and $\mathcal{I}_2 = [U, V_2]$ be two closed intervals in $\mathcal{R}$ with same initial starting point U and non-overlapping endpoints $V_1 \cap V_2 = \varnothing$. A member $\lambda \in \mathcal{R}$ is said to be a *bifurcation* of $\mathcal{I}_1$ and $\mathcal{I}_2$ iff $[U, \lambda] = \mathcal{I}_1 \cap \mathcal{I}_2$. A member $\xi \in \mathcal{R}$ is said to be a *discriminant* of $\mathcal{I}_1$ from $\mathcal{I}_2$ iff $\mathcal{I}_1 \setminus \mathcal{I}_2 = [\xi, V_1]$.

For instance, relative to the causal space in Fig. 8, consider the intervals

$$[S_0, S_7]_{\mathcal{R}} \quad \text{and} \quad [S_0, S_9]_{\mathcal{R}}.$$

Then, the bifurcation is S_1, because

$$[S_0, S_7]_{\mathcal{R}} \cap [S_0, S_9]_{\mathcal{R}} = [S_0, S_1]_{\mathcal{R}};$$

and their discriminants are S_3 and S_4 respectively, because

$$[S_0, S_7]_{\mathcal{R}} \setminus [S_0, S_9]_{\mathcal{R}} = [S_3, S_7]_{\mathcal{R}} \quad \text{and}$$
$$[S_0, S_9]_{\mathcal{R}} \setminus [S_0, S_7]_{\mathcal{R}} = [S_4, S_9]_{\mathcal{R}}.$$

In principle, bifurcations and discriminants might not exist; or if they exist, they might not be unique. The following lemma disproves this possibility by showing that bifurcations and discriminants always exist and are unique.

Lemma 14. *Let $\mathcal{R}$ be a set of realisations, and let $\mathcal{I}_1 = [U, V_1]$ and $\mathcal{I}_2 = [U, V_2]$ be two closed intervals in $\mathcal{R}$ with same initial starting point U and non-overlapping endpoints $V_1 \cap V_2 = \varnothing$. Then, there exists*

a) *a unique bifurcation of $\mathcal{I}_1$ and $\mathcal{I}_2$;*
b) *a unique discriminant of $\mathcal{I}_1$ from $\mathcal{I}_2$;*
c) *and a unique discriminant of $\mathcal{I}_2$ from $\mathcal{I}_1$.*

Proof. First, we observe that there cannot be any $V \in \mathcal{I}_1$ such that $V \cap V_1 = \varnothing$. For if this was true, then we would have that for all $W \in [V, V_1]$, $W \cap V_1 = \varnothing$, which would lead to a contradiction since we know that for $W = V_1$, $W \cap V_1 \neq \varnothing$. Repeating the same argument for $\mathcal{I}_2$, we also conclude that there cannot be any $V \in \mathcal{I}_2$ such that $V \cap V_2 = \varnothing$.

Second, using a similar argument as above, if $V \in \mathcal{I}_1$ is such that $V \cap V_2 = \varnothing$, then for all $W \in [V, V_1]$, $W \cap V_2 = \varnothing$; and if $V \in \mathcal{I}_1$ is such that $V \cap V_2 \neq \varnothing$, then for all $W \in [\Omega, V]$, $W \cap V \neq \varnothing$. This leads us to conclude that $\mathcal{I}_1$ can be partitioned into

$$[U, R_1) := \{W \in \mathcal{I}_1 : W \cap V_2 \neq \varnothing\}$$
$$\text{and} \quad (S_1, V_1] := \{W \in \mathcal{I}_1 : W \cap V_2 = \varnothing\},$$

for some $R_1, S_1 \subset \Omega$, that is to say, where $\mathcal{I}_1 = [U, R_1) \cup (S_1, V_1]$ and $[U, R_1) \cap (S_1, V_1] = \varnothing$. But, due to axiom R4, both intervals must be closed. Hence, in particular, it is true that $[\Omega, R_1) = [\Omega, \lambda_1]$ for some $\lambda_1 \in \mathcal{I}_1$. Similarly, $\mathcal{I}_2$ can be partitioned into

$$[U, R_2) := \{W \in \mathcal{I}_2 : W \cap V_1 \neq \varnothing\}$$
$$\text{and} \quad (S_2, V_2] := \{W \in \mathcal{I}_2 : W \cap V_1 = \varnothing\},$$

and again, $[U, R_2) = [U, \lambda_2]$ for some $\lambda_2 \in \mathcal{I}_2$. Now, if $W \in \mathcal{I}_1 \cup \mathcal{I}_2$, then $W \in [U, \lambda_1] \Leftrightarrow W \in [U, \lambda_2]$. Hence, $\lambda_1 = \lambda_2$ is unique and is a bifurcation, proving part (a). For parts (b) and (c), we note that $(R_1, V_1] = [\xi_1, V_1]$ and $(R_2, V_2] = [\xi_2, V_2]$ for some $\xi_1 \in \mathcal{I}_1 \setminus \mathcal{I}_2$ and $\xi_2 \in \mathcal{I}_2 \setminus \mathcal{I}_1$ respectively due to axiom R4. But since $\mathcal{I}_1 \setminus \mathcal{I}_2 = \mathcal{I}_1 \setminus [U, \lambda_1] = [\xi_1, V_1]$, and similarly $\mathcal{I}_2 \setminus \mathcal{I}_1 = [\xi_2, V_2]$, the members ξ_1 and ξ_2 are the desired discriminants, and they are unique. □

The important consequence of this lemma is that there is always a pair of closed intervals $[\lambda, \xi_1]$ and $[\lambda, \xi_2]$ that *precisely* capture the sub-process, or *mechanism*, during which a realisation can split into two mutually exclusive causal branches.

To intervene a random experiment in order to give rise to a particular event A, we first need to identify all the sub-processes that can split the course of the realisation into intervals leading to A or its negation A^c. These sub-processes will start and end at instants that will be called A-bifurcations and A-discriminants respectively. Again, a lemma guarantees that these sub-processes exist and are unique.

Definition 15 (A-Bifurcations, A-Discriminants)**.** Let $\mathcal{R}$ be a set of realisations, and let $A \in \sigma(\mathcal{R})$ be a member of the generated σ-algebra of $\mathcal{R}$.

1) A member $\lambda \in \mathcal{R}$ is said to be an *A-bifurcation* iff it is a bifurcation of two intervals $[\Omega, V_A]$ and $[\Omega, V_{A^c}]$ with endpoints V_A and V_{A^c} in some representations of A and A^c respectively. The *set of A-bifurcations* is the subset $\lambda(A) \subset \mathcal{R}$ of all A-bifurcations.
2) Let $\lambda \in \lambda(A)$ be an A-bifurcation. A member $\xi \in \mathcal{R}$ is said to be an *A-discriminant for λ* iff there exists V_A in a representation of A such that $[\xi, V_A] = [\Omega, V_A] \setminus [\Omega, \lambda]$. The *set of A-discriminants for λ* is denoted as $\xi(\lambda)$.

Figure 9 illustrates the set of A-bifurcations for A defined as

$$A = S_7 \cup S_9 \cup S_{12} \cup S_{13} = \{1, 3, 4, 5, 6, 7\}.$$

The set of A-bifurcations is

$$\lambda(A) = \{S_0, S_1, S_2, S_3, S_4\}.$$

Each member has an associated set of A-discriminants. For instance,

$$\xi(S_0) = \{S_1, S_2\} \quad \text{and} \quad \xi(S_2) = \{S_5\}.$$

The critical bifurcations that appear in the figure are a subset of the bifurcations, and they will be defined later in the text.

Lemma 16. *Let $\mathcal{R}$ be a set of realisations, and let $A \in \sigma(\mathcal{R})$ be a member of the generated σ-algebra of $\mathcal{R}$. Then, the set*

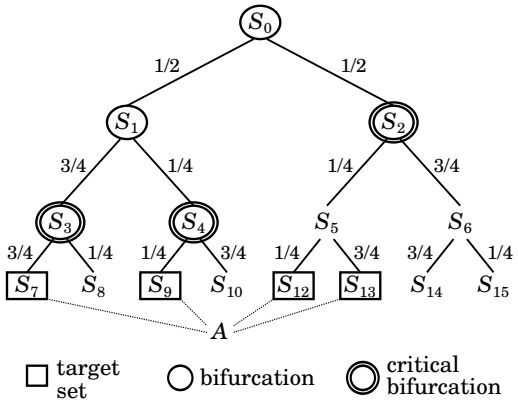


Fig. 9. A-Bifurcations.

of A-bifurcations $\lambda(A)$ and the sets of A-discriminants $\xi(\lambda)$, $\lambda \in \lambda(A)$, exist, are countable and unique.

Proof. Let $\mathcal{A}_1$ and $\mathcal{A}_1^c$ be representations of A and A^c respectively. Consider the set $\lambda_1(A) \subset \mathcal{R}$ of bifurcations of $[\Omega, V_A]$ and $[\Omega, V_{A^c}]$ generated by all pairs of endpoints $(V_A, V_{A^c}) \in \mathcal{A} \times \mathcal{A}^c$. Due to the representation theorem, we know that both $\mathcal{A}$ and $\mathcal{A}^c$ are countable. Therefore, $\mathcal{A}\times\mathcal{A}^c$ and $\lambda_1(A)$ are countable too. Now, repeat the same procedure to construct a set $\lambda_2(A)$ of bifurcations from two representations $\mathcal{A}_2$ and $\mathcal{A}_2^c$ of A and A^c respectively.

Let $R \in \lambda_1(A)$. Then, there exists $R_A \in \mathcal{A}_1$ and $R_{A^c} \in \mathcal{A}_1^c$ such that $[\Omega, R] = [\Omega, R_A] \cap [\Omega, R_{A^c}]$. Since $\mathcal{A}_2$ and $\mathcal{A}_2^c$ are representations of A and A^c respectively, there must be members $S_A \in \mathcal{A}_2$ and $S_{A^c} \in \mathcal{A}_2^c$ such that $R_A \cap S_A \neq \varnothing$ and $R_{A^c} \cap S_{A^c} \neq \varnothing$. Due to axiom R2, it must be that either $R_A \subset S_A$ or $S_A \subset R_A$; similarly, either $R_{A^c} \subset S_{A^c}$ or $S_{A^c} \subset R_{A^c}$. But then, $[\Omega, S_A] \cap [\Omega, S_{A^c}] = [\Omega, R]$, implying that $R \in \lambda_2(A)$. Since R is arbitrary, $\lambda_1(A) = \lambda_2(A)$. Hence, we have proven that $\lambda(A)$ exists, is countable and unique.

Let $\lambda \in \lambda(A)$ be an arbitrary A-bifurcation. Let $\mathcal{A}_1$ and $\mathcal{A}_2$ be two representations of A. Because A-representations are countable, there exists a countable number of intervals $[\Omega, V_A]$, $V_A \in \mathcal{A}_1$, containing λ and an associated discriminant. Let $\xi_1(\lambda)$ be the collection of those discriminants. Following an argument analogous as above, it is easy to see that $\xi_2(\lambda)$, the set of discriminants constructed from the intervals associated to the representation $\mathcal{A}$, must be equal to $\xi_1(\lambda)$. Hence, for each $\lambda \in \lambda(A)$, $\xi(\lambda)$ exists, is countable and unique. □

We are now ready to define interventions on a causal space. In the next definition, an A-intervention is defined as the change of the causal measure at the bifurcations such that the desired event A will inevitably take place. This is done by removing all the probability mass leading to the undesired event A^c and renormalizing thereafter.

Definition 17 (Intervention)**.** Let $\mathcal{R}$ be a set of realisations, P be a causal measure, and A be a member of the generated σ-algebra of $\mathcal{R}$. A causal measure P' is said to be an *A-intervention of* P iff for all $U, V \in \mathcal{R}$ such that $V \cap A \neq \varnothing$,

$$\mathsf{P}'(V|U) \cdot G(U,V) = \mathsf{P}(V|U), \tag{1}$$

where $G(U,V)$ is the *gain* of the interval $[U,V]$ defined by

$$G(U,V) := \prod_{\lambda\in\Lambda} \sum_{\xi\in\xi(\lambda)} \mathsf{P}(\xi|\lambda). \tag{2}$$

Here, $\Lambda := [U,V] \cap \lambda(A)$ is the set of bifurcations in $[U,V]$, and each $\xi(\lambda)$ is the set of A-discriminants of $\lambda \in \Lambda$.

In the tree visualization, an A-intervention can be thought of as a reallocation of the probability mass at the A-bifurcations (see Fig. 10). Essentially, this is done by first removing the probability mass from the transitions that do not have any successor realization in A and then by renormalizing the remaining transitions, i.e. the ones rooted at A-discriminants.

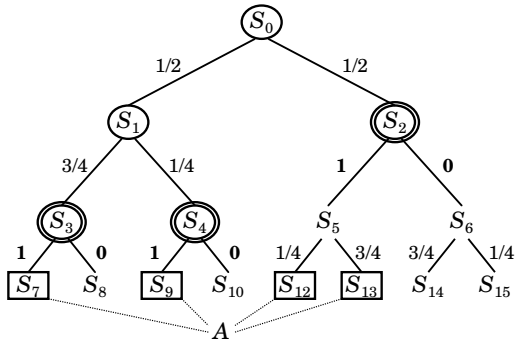


Fig. 10. A-Intervention.

Theorem 18 (Uniqueness of A-Interventions)**.** *Let $\mathcal{R}$ be a set of realisations, P be a causal measure, and A be a member of the generated σ-algebra of $\mathcal{R}$. The A-intervention is unique if for each bifurcation $\lambda \in \lambda(A)$, the corresponding A-discriminants are not null, i.e. they are such that*

$$\sum_{\xi\in\xi(\lambda)} \mathsf{P}(\xi|\lambda) > 0. \tag{3}$$

Proof. Let P' be an A-intervention. Because it is a causal measure, $\mathsf{P}'(V|U) = 1$ when $V \supset U$ and $\mathsf{P}'(V|U) = 0$ when $V \cap U = \varnothing$. It remains to check that $\mathsf{P}'(V|U)$ is unique when $V \subset U$. If $V \cap A \neq \varnothing$, then the definition applies. Here, $\lambda(A)$ and the $\{\xi(\lambda)\}_{\lambda\in\lambda(A)}$ are unique due to Lemma 14, thus so are the $\Lambda := [U,V]\cap\lambda(A)$ for each $U, V \in \mathcal{R}$. Hence, we see that if condition (3) holds for all $\lambda \in \lambda(A)$, then (1) has a unique solution for $\mathsf{P}'(V|U)$. Finally, if $V \cap A = \varnothing$ then $\mathsf{P}'(V|U)$ depends on where $[U,V]$ contains an A-bifurcation. If it does not, then $\mathsf{P}'(V|U) = \mathsf{P}(V|U)$. If it does, then Axiom C3 implies $\mathsf{P}'(V|U) = 0$. □

Corollary 19. *A-interventions are unique up to intervals containing only null discriminants. In other words, given two A-interventions P'_1 and P'_2 let $U, V \in \mathcal{R}$. Then, $\mathsf{P}'_1(V|U) = \mathsf{P}'_2(V|U)$ if for all $\lambda \in [U,V] \cap \lambda(A)$, there exists $\xi \in \xi(\lambda)$ such that $\mathsf{P}(\xi|\lambda) > 0$.*

Finally, the next proposition shows that the intervention is indeed correct in the sense that the desired event occurs with certainty after the intervention.

Proposition 20. *Let $\mathcal{R}$ be a set of realisations and let A be a member of the generated σ-algebra Σ of $\mathcal{R}$. Furthermore, let*

P' *be probability measure compatible with an* A*-intervention of a causal measure* P*. Then,* $\mathsf{P}'(A|\Omega) = 1$.

Proof. Let $\mathcal{A}$ and $\mathcal{A}^c$ be representations of A and A^c in $\mathcal{R}$ respectively. Then, each $V \in \mathcal{A}^c$ is such that $V \cap A = \varnothing$. Hence, due to (I1),

$$\mathsf{P}'(A^c|\Omega) = \sum_{V \in \mathcal{A}^c} \mathsf{P}'(V|\Omega) = 0.$$

But then, since P' is a probability measure,

$$\mathsf{P}'(A|\Omega) = 1 - \mathsf{P}'(A^c|\Omega) = 1.$$

□

A closer look at the definition of an A-intervention reveals that, while the set of bifurcations $\lambda(A)$ contains all the logically required bifurcations (i.e. all moments having a branch leading to the undesired event A^c), some of them remain unaltered after the intervention. In particular, this is always the case when the mechanisms assign zero probability to the branches leading to A^c.

Definition 21 (Critical Bifurcations)**.** Let $\mathcal{R}$ be a set of realisations, P be a causal measure, and A be a member of the generated σ-algebra of $\mathcal{R}$. A bifurcation $\lambda \in \lambda(A)$ is said to be *critical* iff the corresponding A-discriminants are not complete, i.e. they are such that

$$\sum_{\xi \in \xi(\lambda)} \mathsf{P}(\xi|\lambda) < 1. \tag{4}$$

Proposition 22. *The gain* (2) *is equal to*

$$G(U, V) = \prod_{\lambda \in \Gamma} \sum_{\xi \in \xi(\lambda)} \mathsf{P}(\xi|\lambda) \tag{5}$$

where Γ *is the set of critical bifurcations in the interval from* U *to* V*, and each* $\xi(\lambda)$ *is the set of* A*-discriminants of* $\lambda \in \Lambda$.

Proof. Partition Λ into Γ and $\overline{\Gamma} = \Lambda \setminus \Gamma$, where Γ contains only the critical A-bifurcations. Then,

$$\begin{aligned}\prod_{\lambda \in \Lambda} \sum_{\xi \in \xi(\lambda)} \mathsf{P}(\xi|\lambda) &= \left\{ \prod_{\lambda \in \overline{\Gamma}} \sum_{\xi \in \xi(\lambda)} \mathsf{P}(\xi|\lambda) \right\} \cdot \left\{ \prod_{\lambda \in \Gamma} \sum_{\xi \in \xi(\lambda)} \mathsf{P}(\xi|\lambda) \right\} \\ &= 1 \cdot \left\{ \prod_{\lambda \in \Gamma} \sum_{\xi \in \xi(\lambda)} \mathsf{P}(\xi|\lambda) \right\} = \prod_{\lambda \in \Gamma} \sum_{\xi \in \xi(\lambda)} \mathsf{P}(\xi|\lambda).\end{aligned}$$

□

C. Random Variables

We recall the formal definition of a random variable. Let $(\Omega, \Sigma, \mathsf{P})$ be a probability space and (S, Ξ) be a measurable space. Then an (S, Ξ)-valued random variable is a function $X : \Omega \to S$ which is (Σ, Ξ)-measurable, i.e. for every member $B \in \Xi$, its preimage $X^{-1}(B) \in \Sigma$ where $X^{-1}(B) = \{\omega : X(\omega) \in B\}$. If we have a collection $(X_\gamma : \gamma \in \Gamma)$ of mappings $X_\gamma : \Omega \to S_\gamma$, then

$$\Sigma := \sigma(X_\gamma : \gamma \in \Gamma)$$

is the smallest σ-algebra Σ such that each X_γ is Σ-measurable.

The lesson we have learnt so far is that it is not enough to just specify the σ-algebra of a random experiment in order to understand the effect of interventions; rather, we need to endow the σ-algebra with a causal structure. While one would expect the same to hold for random variables one wishes to intervene, we will see that it is not necessary to explicitly model causal dependencies among them. Instead, it is sufficient to establish a link to some abstract causal space that is shared by all the random variables. We begin this investigation thus with a definition of a function having causal structure.

Definition 23 (Realisable Function)**.** Let Ω, S be sets and $\mathcal{R}$ and $\mathcal{S}$ be sets of realisations over Ω and S respectively. A function $X : \Omega \to S$ is said to be $(\mathcal{R}, \mathcal{S})$-realisable iff for every $B \in \mathcal{S}$, the preimage $X^{-1}(B)$ is a member of $\mathcal{R}$. The following picture illustrates this:

$$\Omega \xrightarrow{X} S$$

$$\mathcal{R} \xleftarrow{X^{-1}} \mathcal{S}$$

The next proposition shows that realisable functions are measurable functions, but not vice versa—as intuition immediately predicts.

Proposition 24. *Let* $\mathcal{R}$ *and* $\mathcal{S}$ *be two sets of realisations over sets* Ω *and* S *respectively. Let* $\Sigma = \sigma(\mathcal{R})$ *and* $\Xi = \sigma(\mathcal{S})$*. If a mapping* $X : \Omega \to S$ *is* $(\mathcal{R}, \mathcal{S})$*-realisable then it is also* (Σ, Ξ)*-measurable. However, the converse is not necessarily true. In a diagram:*

$$\begin{array}{ccc} \Omega \xrightarrow{X} S & & \Omega \xrightarrow{X} S \\ & \Longrightarrow & \\ \mathcal{R} \xleftarrow[X^{-1}]{} \mathcal{S} & & \Sigma \xleftarrow[X^{-1}]{} \Xi \end{array}$$

Proof. Let $B \in \Xi$. Then, Theorem 4 tells us that there exists a representation $\mathcal{B}$ of B in $\mathcal{S}$. Since X is $(\mathcal{R}, \mathcal{S})$-realisable, every member $V \in \mathcal{B}$ has a preimage that is in $\mathcal{R}$, i.e. $X^{-1}(V) \in \mathcal{R}$. But $\bigcup_{V \in \mathcal{B}} X^{-1}(V) = X^{-1}(B)$ and $\bigcup_{V \in \mathcal{B}} X^{-1}(V) \in \Sigma$, hence X is (Σ, Ξ)-measurable.

To disprove the converse, consider the following counterexample. Take $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$ and $S = \{s_1, s_2\}$. Let $\mathcal{R} = \{R_1, \ldots R_8\}$, where: $R_1 = \Omega$; $R_2 = \{\omega_1, \omega_2\}$; $R_3 = \{\omega_3, \omega_4\}$; $R_4 = \{\omega_1\}$; $R_5 = \{\omega_2\}$; $R_6 = \{\omega_3\}$; $R_7 = \{\omega_4\}$; and $R_8 = \varnothing$. Furthermore, let $\mathcal{S} = \{S_1, S_2, S_3, S_4\}$, where: $S_1 = S$; $S_2 = \{s_1\}$; $S_3 = \{s_2\}$; and $S_4 = \varnothing$. Let X be such that $X(R_2) = S_2$ and $X(R_6) = S_3$. Observe that $S_1 = S_2 \cup S_3 \in \Xi$ and $S_4 = S_1^c \in \Xi$, and in particular $S_1, S_4 \in \mathcal{S}$. However, $X^{-1}(S_1) = R_2 \cup R_6$, which is obviously in Σ but not in $\mathcal{R}$. □

Next, realisable random variables are simply defined as realisable functions endowed with a causal measure.

Definition 25 (Realisable Random Variable)**.** Let $(\Omega, \mathcal{R}, \mathsf{P})$ be a causal space. A *realisable random variable* X is an (S, Ξ)-valued function that is $(\mathcal{R}, \mathcal{S})$-realisable, where $\Xi = \sigma(\mathcal{S})$.

Finally, we define the intervention of a random variable. Let B be a measurable event in Ξ, the σ-algebra in the range of X. Then, a B-intervention of X is done in two steps: first, B is translated into its corresponding event A living in the abstract σ-algebra Σ; second, the resulting event A is intervened.

Definition 26 (Intervention of a Random Variable)**.** Let $(\Omega, \mathcal{R}, \mathsf{P})$ be a causal space and let X be a $(\mathcal{R}, \mathcal{S})$-realisable random variable. Given a set $B \in \Xi = \sigma(\mathcal{S})$ of the generated σ-algebra, a *B-intervention of the realisable random variable* X is a $X^{-1}(B)$-intervention of the causal measure P.

Corollary 27. *B-interventions of a realisable random variable are unique up to intervals containing only null discriminants.*

This concludes our abstract model of causality. It is immediately seen that a causal stochastic process can be characterized as a collection $(X_\gamma : \gamma \in \Gamma)$ of $(\mathcal{R}, \mathcal{S}_\gamma)$-realisable random variables $X : \Omega \to S_\gamma$ respectively defined over a shared causal space $(\Omega, \mathcal{R}, \mathsf{P})$. This is in perfect accordance with the theory so far developed.

VI. Discussion

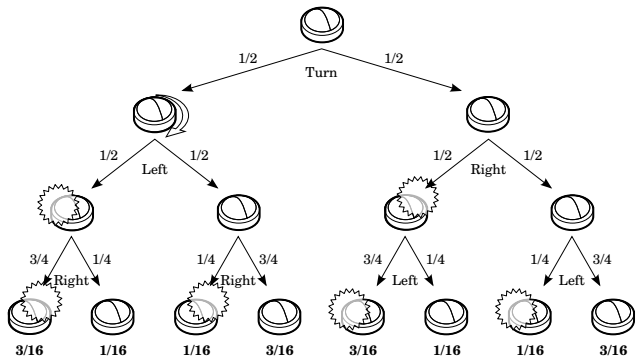
A. Causal Spaces

Rather than replacing existing frameworks, the model of the subject presented in the previous section aims at providing a common ground containing basic elements to define causal interventions. This was achieved by supplying the σ-algebra of standard probability theory with a set of realisations encoding the causal dependencies in its events.

The model is limited to countable sets of realisations. This was chosen so as to guarantee that the σ-algebras generated by realisation sets are always well defined (e.g. do not have excessive cardinalities). Also, notice that the definition of causal interventions critically depends on having closed realisation intervals to uniquely determine the bifurcation points of causal histories.

On the other hand, the model can accommodate very general causal dependencies. For instance, consider the causal induction problem taken from Ortega (2011) and Ortega and Braun (2014). Here we have an electronic game with two indistinguishable panels that, at a prescribed time instant, light up (simultaneously) with a fixed probability. Internally, the electronics are wired such that the state of one light (master) influences the state of the other (slave), but not the other way around. The player can, if she so wishes, set the state of any of the two lights by tapping the respective panel: say, a quick (prolonged) touch switches the light on (off). The electronic device is initially placed on the table with an unknown orientation, with one panel on the left and the other on the right. The aim is to figure out which of the two is the master panel.

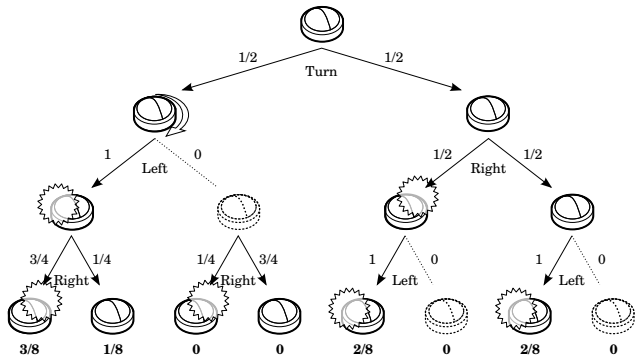
The corresponding causal model is depicted in Fig. 11. The root node (the Turn variable) represents the unknown hypothesis, and each branch leads to one of the two causal hypotheses over the panels. Crucially, note that the two hypotheses are observationally indistinguishable: the player must intervene to render them identifiable—for instance, by turning on the light



Left	Right	P(Left, Right | Turn = yes)	P(Left, Right | Turn = no)
no	no	3/8	3/8
no	yes	1/8	1/8
yes	no	1/8	1/8
yes	yes	3/8	3/8

Fig. 11. An Electronic Game of Causal Induction.

of the left panel (Fig. 12). The causal dependencies in this simple induction example *cannot* be captured using a causal directed acyclic graph.



Left	Right	P′(Left, Right | Turn = yes)	P′(Left, Right | Turn = no)
no	no	3/4	1/2
no	yes	1/4	0
yes	no	1/4	1/2
yes	yes	3/4	0

Fig. 12. A causal intervention breaks the symmetry of the causal hypotheses.

B. What is an Action?

Classical decision theory assumes from the outset that there is a clear-cut distinction between action and observation variables (Von Neumann and Morgenstern, 1944; Savage, 1954). Similarly, control theory and artificial intelligence take this distinction as a given. Indeed, artificial intelligence textbooks describe an agent as any system possessing sensors and effectors (Russell and Norvig, 2009).

In contrast, the idea of the subject as a construct plus the line of thought developed in this paper suggest a different story: the distinction between actions and observations is *inferred* rather

than given. More specifically, given a starting hypothesis, one can generate new hypotheses in two ways: through the application of causal interventions to change the role of random variables from observations to actions *and* through the application of inverse causal interventions to achieve the opposite. *Accordingly, a random variable is identified as an action if and only if it can be thought of as the result of a causal intervention*. In this sense, causal interventions merely add a convenient syntax to relate hypotheses in Bayesian probability theory. If a random variable is unambiguously classified as either an action or an observation, then it is only because the agent does not possess hypotheses offering competing explanations. In particular, the setting in classical decision theory is just a special case.

This also sheds more light into the connection between causal interventions and Lacan's *objet petit a*[4]. An action turns out to be a random variable of somewhat contradictory nature: because on one hand, it is statistically independent and hence, subjectivised; but on the other hand, it is still generated by the external world, namely, by the very decision processes of the subject that are not under its direct control, e.g. following a utility-maximizing agenda. Lacan's term *objet petit a* can thus be regarded as a play of words that encapsulates this dual nature (Fink, 1996).

C. Concluding Remarks

There are numerous reasons why I chose to compare Bayesian probability theory to Lacanian theory. It is true that, virtually since its inception, psychoanalytic theories have always faced fierce opposition that has questioned their status as a scientific discipline (see e.g. Popper, 1934; Feynman et al., 1964). While their efficacy as a treatment of mental illnesses is undoubtedly controversial (Tallis, 1996), cultural studies have embraced psychoanalytic theories as effective conceptual frameworks to structure the discourse about subjectivity that is metaphysically frugal (Mansfield, 2000). As a researcher in artificial intelligence, the greatest value I see in the psychoanalytic theories is in that they epitomise the contingent cultural assumptions about subjectivity of modern Western thinking, summarizing ideas that otherwise would require a prohibitive literature research. Finally, it must also be pointed out that the abstract model of the subject presented here does not constitute a scientific theory in the Popperian sense; only instantiated causal spaces, given appropriate experimental conditions, can be falsified.

VII. Acknowledgements

The writing of this essay has benefited from numerous conversations with D.A. Braun, D. Balduzzi, J.R. Donoso, A. Saulton and E. Wong. Furthermore, I wish to thank M. Hutter and J.P. Cunningham for their comments on a previous version of the abstract model of causality, and A. Jori, Z. Domotor and A.P. Dawid for their insightful lectures and discussions on ancient philosophy, philosophy of science and statistical causality at the Universities of Tübingen and Pennsylvania. This study was funded by the Ministerio de Planificación de Chile (MIDEPLAN); the Emmy Noether Grant BR 4164/1-1 (Computational and Biological Principles of Sensorimotor Learning); and by grants from the U.S. National Science Foundation, Office of Naval Research and Department of Transportation.

[4]The term *objet petit a* loosely translates into "object little other". The "*a*" in *objet petit a* stands for the French word *autre* (other).

References

Ash, R. and Doléans-Dade, C. (1999). *Probability & Measure Theory*. Academic Press, 2nd edition.

Bayes, T. (1763). An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philosophical Transactions*, 53:370–418.

Billingsley, P. (1978). *Ergodic theory and information*. R. E. Krieger Pub. Co.

Burkitt, I. (2008). Subjectivity, Self and Everyday Life in Contemporary Capitalism. *Subjectivity*, 23:236–245.

Cantor, G. (1874). Ueber eine Eigenschaft des Inbegriffes aller reellen algebraischen Zahlen. *Journal für die reine und angewandte Mathematik*, 77:258–262.

Cartwright, N. (1983). *How the Laws of Physics Lie*. Oxford University Press.

Chomsky, N. (1957). *Syntactic Structures*. Mouton & Co.

Cox, R. (1961). *The Algebra of Probable Inference*. Johns Hopkins.

Dawid, A. (2014). Statistical Causality from a Decision-Theoretic Perspective. *arXiv:1405.2292*.

Dawid, A. P. (2007). Fundamentals of Statistical Causality. Research Report 279, Department of Statistical Science, University College London.

Dawid, P. (2010). Seeing and Doing: The Pearlian Synthesis. In Dechter, R., Geffner, H., and Halpern, J., editors, *Heuristics, Probability and Causality*. Cambridge Univesity Press.

De Finetti, B. (1937). La Prévision: Ses Lois Logiques, Ses Sources Subjectives. In *Annales de l'Institut Henri Poincaré*, volume 7, pages 1–68.

de Saussure, F. (1916). *Cours de linguistique générale (Course in General Linguistics)*.

Descartes, R. (1637). *Discours de la méthode: pour bien conduire sa raison, et chercher la vérité dans les sciences (Discourse on the Method)*.

Feynman, R., Leighton, R., and Sands, M. (1964). *The Feynman Lectures on Physics (Vol. I)*. Adison-Wesley.

Fink, B. (1996). *The Lacanian Subject* . Princeton University Press.

Foucault, M. (1964). *Histoire de la folie à l'âge classique (Madness and civilization: A history of insanity in the age of reason)*. Pantheon Books.

Foucault, M. (1966). *Les Mots et les choses (The Order of Things: An Archaeology of the Human Sciences)*. Éditions Gallimard.

Frege, G. (1892). Über Sinn und Bedeutung (On Sense and Reference). *Zeitschrift für Philosophie und Philosophische Kritik*, 100:25–50.

Freud, S. (1899). *Die Traumdeutung (The Interpretation of Dreams)*. Franz Deuticke, Leipzig & Vienna.

Freud, S. (1920). *Jenseits des Lustprinzips (Beyond the Pleasure Principle)*. Internationaler Psychoanalytischer Verlag, G.M.B.H.

Hegel, G. (1807). *Phänomenologie des Geistes (Phenomenology of Spirit)*.

Heidegger, M. (1927). *Sein und Zeit*. Max Niemeyer, Tübingen.

Hume, D. (1748). *An Enquiry Concerning Human Understanding*.

Jaynes, E. and Bretthorst, L. (2003). *Probability Theory: The Logic of Science: Books*. Cambridge University Press.

Kim, J. (2005). *Philosophy of Mind* . Dimensions of Philosophy. Westview Press, 2nd edition edition.

Knape, J. (2000). *Was ist Rhetorik? (What is Rhetoric)*. Reclam, Philipp, jun. GmbH.

Kolmogorov, A. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Springer, Berlin.

Lacan, J. (1973). *Les quatre concepts fondamentaux de la psychanalyse (The Four Fundamental Concepts of Psychoanalysis)*. Le Seuil.

Lacan, J. (1977). *Fonction et champ de la parole et du langage en psychanalyse (The function and field of speech and language in psychoanalysis)*, pages 30–113. Tavistock Publications. Translated by Alan Sheridan.

Laplace, P. S. (1774). Mémoires sur la probabilité des causes par les évènements. *Mémoires de mathématique et des physiques presenteés à l'Académie royale des sciences, par divers savans, & lûs dans ses assemblées*, 6:621–656.

Lauritzen, S. L. and Richardson, T. S. (2002). Chain graph models and their causal interpretations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(3):321–348.

Lebesgue, H. (1902). *Integrale, longueur, aire*. PhD thesis, Université de Paris.

Mansfield, N. (2000). *Subjectivity: Theories of the Self from Freud to Haraway*. NYU Press.

Maturana, H. (1970). Biology of Cognition. Technical report, Biological Computer Laboratory: Urbana, IL.

Maturana, H. and Varela, F. (1987). *The Tree of Knowledge: The biological roots of human understanding*. Shambala Publications, Inc, 1st edition.

Nietzsche, F. (1887). *Zur Genealogie der Moral (On the Genealogy of Morality)*.

Ortega, P. (2011). Bayesian Causal Induction. In *2011 NIPS Workshop in Philosophy and Machine Learning*.

Ortega, P. and Braun, D. (2014). Generalized Thompson Sampling for Sequential Decision-Making and Causal Inference. *Complex Adaptive Systems Modeling*, 2(2).

Osborne, M. and Rubinstein, A. (1999). *A Course in Game Theory*. MIT Press.

Pearl, J. (1993). Graphical Models, Causality, and Intervention. *Statistical Science*, 8(3):266–273.

Pearl, J. (2009). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge, UK.

Pearson, K., Lee, A., and Bramley-Moore, L. (1899). Mathematical Contributions to the Theory of Evolution. VI. Genetic (Reproductive) Selection: Inheritance of Fertility in Man, and of Fecundity in Thoroughbred Racehorses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 192:257–330.

Popper, K. (1934). *The Logic of Scientific Discovery (Routledge Classics)*. Routledge.

Quine, W. (1951). Two Dogmas of Empiricism. *The Philosophical Review*, 60(1):20–43. Reprinted in W.V.O. Quine, From a Logical Point of View (Harvard University Press, 1953; second, revised, edition 1961.

Ramsey, F. P. (1931). *The Foundations of Mathematics and Other Logical Essays*, chapter 'Truth and Probability'. Harcourt, Brace and Co., New York, reprinted edition.

Rubin, D. (1974). Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies. *Journal of Educational Psychology*, 66(5):688–701.

Russell, B. (1905). On Denoting. *Mind*, 14:479–493.

Russell, B. (1913). On the Notion of Cause. *Proceedings of the Aristotelian Society*, 13:1–26.

Russell, S. and Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*. Prentice-Hall, Englewood Cliffs, NJ, 3rd edition edition.

Salmon, W. (1980). Probabilistic Causality. *The Pacific Philosophical Quarterly and Personalist*, 61(1–2):50–74.

Savage, L. (1954). *The Foundations of Statistics*. John Wiley and Sons, New York.

Sejnowski, T. and van Hemmen, J. (2006). *23 Problems in Systems Neuroscience*. Oxford University Press.

Shafer, G. (1996). *The Art of Causal Conjecture*. MIT Press.

Simon, H. (1977). *Models of Discovery: and Other Topics in the methods of Science*. Boston Studies in the Philosophy and History of Science (Book 54). Springer.

Spirtes, P. and Scheines, R. (2001). *Causation, Prediction, and Search, Second Edition*. MIT Press.

Suppes, P. (1970). *A Probabilistic Theory of Causality*. Amsterdam: North-Holland Publishing Company.

Tallis, R. (1996). Burying Freud. *Lancet*, 347(9002):669–671.

Turing, A. M. (1936–1937). On computable numbers, with an application to the *Entscheidungsproblem*. *Proceeding of the London Mathematical Society, series 2*, 42:230–265.

Von Neumann, J. and Morgenstern, O. (1944). *Theory of Games and Economic Behavior*. Princeton University Press, Princeton.

Žižek, S. (1992). *Looking Awry: An Introduction to Jacques Lacan through Popular Culture*. The MIT Press.

Žižek, S. (2009). *The sublime object of ideology*. The Essential Žižek. Verso, 2nd edition.

Williams, D. (1991). *Probability with Martingales*. Cambridge mathematical textbooks. Cambridge University Press.

Wittgenstein, L. (1921-1933). *Logisch-Philosophische Abhandlungen (Tractatus Logico-Philosophicus)*. Annalen der Naturphilosophie.

Wittgenstein, L. (1953). *Philosophische Untersuchungen (Philosophical Investigations)*. Blackwell. Translated by G.E.M. Anscombe.

Woodward, J. (2013). Causation and Manipulability. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*.

Winter 2013 edition.