A BAYESIAN REDUCTION OF CAUSATION

DANIEL A. HERRMANN, AYDIN MOHSENI,
BENJAMIN A. LEVINSTEIN, AND BRUCE RUSHING

Abstract. De Finetti conjectured that causation, like chance, reduces to patterns in subjective probability. Recent results on the Pearl Causal Hierarchy seem to foreclose this possibility, proving that causal claims are irreducible to probabilistic ones within structural causal models (SCMs). We show that the reduction succeeds after all. The apparent irreducibility arises from an implicit restriction on the agent's representational resources: once the variable set is enriched to include naturalized intervention variables, interventional propositions reduce to ordinary probabilistic conditioning. To complete the subjectivist picture, we prove a de Finetti-style representation theorem establishing that an agent with exchangeable credences can be represented as simultaneously learning causal structure and the probabilities within that structure via standard Bayesian updating. We then show how this framework bears on the debate between evidential and causal decision theory, arguing that causal decision theory functions as a useful approximation of evidential reasoning over the full algebra. Within SCMs, causal reasoning reduces to subjective probability.

The notion of "cause" thus depends on the notion of probability, and it follows also from the same subjective source as do all judgments of probability: this explanation seems to constitute the true logical translation of the conception of "cause" advanced by David Hume, which I consider the highest peak that has been reached by philosophy. The subjectivist theory of probability will thus be able to open the field of science to this conception, whose significance and value seem not to have been sufficiently understood nor appreciated until now.

—de Finetti (1937) *Foresight: Its Logical Laws, its Subjective Sources* p. 155

1. Introduction

Bayesian reasoning stands as one of the most successful frameworks for rational inference. Representation theorems establish that agents satisfying minimal coherence constraints must have probabilistic credences. Accuracy arguments show that probabilistic beliefs uniquely avoid being accuracy dominated. Dutch-book arguments demonstrate that violating probabilistic coherence leads to exploitable irrationality. These results suggest that probability theory provides the correct normative framework for reasoning under uncertainty across a remarkably broad range of contexts.[1]

[1]For representation theorems see Savage (1954) and Jeffrey (1983). Fishburn (1981) gives an excellent overview of a wide range of representation theorems. For accuracy arguments see Joyce (1998) and Pettigrew (2016). For Dutch-book arguments see Ramsey (1931) and Pettigrew (2020).

In addition to articulating a detailed vision of inference, Bayesianism also sheds light on various puzzling concepts with apparent metaphysical commitments. In particular, Bruno de Finetti's subjective Bayesian program shows we can understand *chance* as patterns in subjective probability. His celebrated representation theorem demonstrates that objective chances can be recovered as artifacts of *exchangeable* credences, that is, credences that are insensitive to the order in which observations occur. An agent who treats coin tosses as exchangeable behaves *as if* she believes in an objective chance about which she is learning, even if chance is absent from her ontology. This approach dissolves metaphysical puzzles about chance while preserving its pragmatic utility.

De Finetti recognized that his subjectivist program should extend to causation. As he wrote in 1937, the subjectivist theory of probability "will thus be able to open the field of science" to a Humean conception of causation as grounded in our credences rather than the world. If successful, such an extension would bring causal reasoning fully within the Bayesian fold.

Yet recent work has appeared to foreclose this possibility for one philosophically and scientifically influential articulation of causation: structural causal models (SCMs). Pearl and Mackenzie (2018) advances the thesis of a causal hierarchy, distinguishing non-causal *associational* reasoning[2] from both *interventional* [3] and *counterfactual* reasoning.[4] Bareinboim et al. (2022) formalize this hierarchy and prove that these levels are strictly irreducible to one another. On this view, de Finetti's project cannot be completed: causation involves something genuinely beyond probability theory.

We argue otherwise. To see the basic idea, consider Stavros, who notices a robust correlation: when "STEAK" appears on a restaurant menu and he orders it, steak typically arrives. One day at a restaurant without "STEAK" on the menu, Stavros pulls out a pen, writes "STEAK" on the menu himself, and orders confidently. The waiter looks confused. "Sir, this is a vegan restaurant."

The standard diagnosis is that Stavros confused *correlation* with *causation*. He failed to recognize that the menu listing and the restaurant's ability to serve steak share a common cause—the chef's decision—and lacked the causal understanding to predict his intervention would fail. But there is another interpretation. Stavros had not been learning about the probability of steak conditional on "STEAK" appearing on the menu in general. He had been learning conditional on it appearing as written by the restaurant. When Stavros wrote it himself, a different event occurred. These are distinct events carrying different conditional probabilities: P(Steak Delivered | STEAK On Menu By Restaurant) is high while P(Steak Delivered | STEAK On Menu By Stavros) is low. If Stavros had tracked this finer partition, pure associational learning would have revealed his mistake. No special causal machinery required.

This vignette illustrates our central thesis: what appears to be a fundamental limitation of probabilistic reasoning actually reflects an incomplete specification of

[2]This is reasoning expressible in the language of conditional probability, sometimes evocatively described as capturing *seeing*.

[3]This reasoning requires a mathematical tool known as the *`do`-operator*, and is sometimes described as capturing *doing*.

[4]This form of reasoning requires still richer machinery, and is sometimes described as capturing *imagining*.

the probability space. When we enrich our conceptual space to distinguish different ways events can occur, interventional reasoning reduces to standard probability theory. The necessity of this enrichment is itself illuminating: it reveals that causal reasoning, within the framework of structural causal models, amounts to associational reasoning over a more finely specified domain. Our reduction thus completes de Finetti's project for a core way of expressing causation, extending the methodology of pragmatic reduction from chance to causation.

Like de Finetti's reduction of chance, our reduction does not fully settle the metaphysics. But it does remove a powerful reason for treating causal structure as an irreducible feature of the world: its alleged indispensability to explanation, learning, and rational action.

A note on the scope of our contribution. We do not aim to reduce all notions of causation. Our reduction applies specifically to structural causal models (SCMs), the dominant framework for causal reasoning in statistics, machine learning, and the social sciences. We do not address geometric and topological notions of causation in spacetime theories, or folk-psychological causal attributions. Nor do we address richer counterfactual theories that appeal to similarity metrics over possible worlds (Halpern 2013; Bareinboim et al. 2022); our reduction concerns only the counterfactual reasoning internal to SCMs, represented as conditioning on intervention variables. However, showing that causation within this influential framework reduces to probability has significant implications for understanding the scope of Bayesian reasoning.[5]

The paper proceeds as follows. §2 explains the pragmatic approach to reduction developed by de Finetti and Brian Skyrms, using their treatment of chance as a template for our treatment of causation. §3 introduces structural causal models. § 4 presents the case against reduction in detail. §5 demonstrates how to naturalize interventions within an enriched probability space. § 6 proves a de Finetti-style representation theorem for causal learning. §7 responds to some objections. These sections establish the reduction at the level of epistemology: causal learning reduces to Bayesian updating over a richer set of variables. But if causal structure remained indispensable to rational agency even after being shown reducible in epistemology, the reduction would be incomplete. §8 explores how our framework bears on the debate between evidential and causal decision theory. Drawing on insights from Meek and Glymour (1994) and others, we develop a picture on which causal decision theory functions as a useful approximation of evidential reasoning over the full algebra, and we identify the choice points at which one might depart from this picture. §9 concludes.

2. Pragmatic Reduction

Our reduction is *pragmatic* rather than definitional, in the spirit of the de Finetti-Skyrms reduction of chance to credence (de Finetti (1937), Skyrms (1984)). This approach to reduction, articulated by Skyrms in his 1984 book *Pragmatics*

[5]Furthermore, there *are* connections between causation as expressed in SCMs and other notions of causation. See Galles and Pearl (1998), Briggs (2012), and Zhang et al. (2013) for discussions of the connections. Though the frameworks don't map perfectly onto each other, given the similarities, we conjecture that a Bayesian reduction is possible for other frameworks. This is fertile ground for future work.

and Empiricism, does not provide definitions or symbolic translations.[6] Rather, it identifies the functional *role* a concept plays in our practices and shows how that role can be fulfilled within a more ontologically austere framework.

To see how this works, consider de Finetti's treatment of chance. Objective chances appear central to scientific reasoning yet seem to go beyond the subjective Bayesian framework. De Finetti's theorem makes room for removing chance from one's ontology without sacrificing its pragmatic value. As Skyrms writes,

> De Finetti is the kind of positivist who doesn't believe in chance—who regards the whole idea as metaphysical excess baggage—but still wants to give an account of the kind of Bayesian reasoning [involved in learning about the chances]. He gives such an account by proving a famous *representation* theorem. In essence, this shows that *one who has degrees of belief which exhibit a certain symmetry behaves as if he believes in chances and is uncertain as to what the correct chance distribution is.* For de Finetti, this demonstrates that belief in the reality of chances is a difference that makes no difference; chances are, for him, simply an artifact of the representation theorem. (pp. 12-13, 1984)

De Finetti's representation theorem does not *force* eliminativism about chance; some draw the opposite lesson, that one should have exchangeable beliefs precisely when one believes in an underlying chance distribution about which one is learning. However, the theorem makes room for the eliminativist view defended by de Finetti and Skyrms. On this view, we can dispense with chances and the puzzles they generate. For instance, we need not independently justify Lewis's Principal Principle (1980): deference to chance emerges from the representation theorem itself.[7] The theorem thus applies pressure on realism about chances by showing we can retain all the pragmatic benefits of chance-talk while avoiding ontological commitment.

Our reduction of causation is similarly pragmatic. The theorems we prove do not *force* eliminativism about causal relations. One could accept our formal results while maintaining that causation is a genuine feature of the world. However, our results show how causal reasoning functions within Bayesian inference without requiring irreducible causal facts. Just as the de Finetti reduction sidesteps metaphysical puzzles about the nature of chance and its connection to credence, our reduction sidesteps parallel puzzles about causation.[8] We can treat causal structure as part of our cognitive apparatus rather than part of the world's fundamental structure. Even those who maintain causal realism should find the reduction informative: it delineates what their realism commits them to beyond probabilistic

[6]Indeed, Skyrms offers criticism of attempts by van Fraassen (1977) and Kyburg (1978) to reduce chance in these other ways.

[7]The Principal Principle plays notoriously poorly with Humean Supervenience. See Briggs (2009), Arntzenius and Hall (2003), Levinstein and Spencer (2025), and Eagle (2025) for discussion.

[8]For example, if SCMs give the right notion of causation, and they are irreducible to probability and other facts, why don't they appear in our best theories of physics? In particular, these theories tend to violate the time asymmetry that is essential to the characterization of causes preceding their effects, which is sometimes used in causal inference in the context of SCMs. In contrast, on our view, causation arises as part of our broader inferential practices, not something in the world itself.

structure, and reveals that this surplus is empirically idle within the SCM framework.[9]

Our reduction begins by endogenizing the notion of an *intervention*. Building on Spirtes et al. (1993), we demonstrate a general method to extend any SCM to an enriched model where interventions are represented within the model's structure, allowing standard probabilistic conditioning to capture information that was represented by interventions in the original model. Within this framework, the notion of intervention—and hence of cause—reduces to statistical association.

We then provide a subjective Bayesian interpretation of these probabilities via a de Finetti-style representation theorem. This theorem identifies patterns of exchangeability in an agent's credences that make her representable as simultaneously learning about causal structure and the probabilities within that structure.

Our formal results are individually straightforward. However, taken together, they allow us to articulate a thoroughly naturalistic and subjectivist account of SCM-style causal reasoning. Like chance in the de Finetti framework, SCMs earn their keep not by representing objective properties of the world, but instead by enabling effective inference.

3. Structural Causal Models

Structural causal models are the dominant framework for causal reasoning across a wide range of disciplines. In statistics and epidemiology, they underwrite methods for identifying causal effects from observational data. In machine learning, they provide the theoretical foundation for causal discovery algorithms and for reasoning about the effects of interventions in complex systems. In the social sciences, they formalize the assumptions needed to draw causal conclusions from non-experimental data. And in philosophy, they serve as the formal backbone of influential accounts of causation (Pearl 2000; Woodward 2005), as well as the basis for interventionist approaches to decision theory, sometimes called *causal* or *interventional* decision theory (Hitchcock 2016; Stern 2019).[10] Their reach is not incidental: SCMs provide what is arguably the most general and widely adopted formal language for expressing causal claims. Showing that reasoning within this framework reduces to probabilistic reasoning is therefore not a narrow technical point but a result with broad implications for how we understand the role of causation in science and rational agency.

We adopt the framework of structural causal models (SCMs) first developed by Spirtes et al. (1993) and then Pearl (2000). An SCM $M = \langle U, V, F, P(U)\rangle$ consists of exogenous variables U, endogenous variables V, structural equations F where each $V_i = f_i(\mathrm{Pa}(V_i), U_i)$ is determined by its endogenous parents and an associated exogenous variable, and a distribution $P(U)$ over the exogenous variables. Following Bareinboim et al. (2022), we restrict attention to SCMs satisfying

[9] Our reduction is in the spirit of those programs that have aimed to show that certain surplus structure in scientific theories are artifacts of the representations provided by those theories (Hertz 1899).

[10] Gallow (2024) discusses how this interventionist picture of causal decision theory relates to other approaches to causal decision theory such as Joyce (1999).

acyclicity,[11] the *causal Markov condition*,[12] *faithfulness*[13] and *mutual independence of exogenous variables*.[14] Under acyclicity and mutual independence of exogenous variables, the structural equations define a deterministic mapping from U to V, so $P(U)$ propagates through the structural equations to induce a unique joint distribution $Q(V)$ over the endogenous variables V.

A simple example makes the framework concrete. Let S (smoking), G (a genetic mutation), and C (cancer) be binary endogenous variables. Consider the SCM M_1 with exogenous variables $U = \{U_S, U_G, U_C\}$, endogenous variables $V = \{S, G, C\}$, structural equations $S = U_S$, $G = f_G(S, U_G)$, $C = f_C(G, U_C)$, and distribution $P(U) = P(U_S)P(U_G)P(U_C)$. The structural equations encode the directed graph $S \to G \to C$. By the causal Markov condition, since the only path from S to C passes through G, the model entails $S \perp\!\!\!\perp C \mid G$.

In an SCM, the structural equations define the causal relationships. A variable X is a *direct cause* of Y when X appears as an argument in Y's structural equation—equivalently, when there is a directed edge from X to Y in the associated graph. X is a *cause* of Y when there is a directed path from X to Y. In M_1, S is a direct cause of G, G is a direct cause of C, and S is a cause of C.

The do-operator computes the distributional consequences of fixing a variable to a particular value. The operation $\texttt{do}(S = 1)$ produces a modified model in which the equation $S = U_S$ is replaced by the constant $S = 1$, while all other equations remain intact. This "graph surgery" removes the dependence of S on U_S while preserving all other structural equations. The resulting interventional distribution $P(C \mid \texttt{do}(S = 1))$ is the distribution over C induced by this modified model.

Our concern is whether causal reasoning within this framework reduces to probabilistic reasoning.[15]

4. The Case Against Reduction

Before presenting our reduction, we examine the case against it. Recent work by Bareinboim et al. (2022) formalizes Pearl's causal hierarchy and proves an irreducibility result that many take to show causal reasoning cannot be captured by purely probabilistic means. Understanding this result is valuable for two reasons. First, it clarifies what any successful reduction must accomplish, and in particular reveals why enriching the agent's representational resources is necessary. Second, given how widely the result has been taken to foreclose reduction, understanding its actual scope is necessary for assessing whether reduction remains viable.

[11]The structural equations admit a recursive ordering of the endogenous variables; equivalently, the associated directed graph has no directed cycles.

[12]Each variable is conditionally independent of its non-descendants given its parents.

[13]Every conditional independence in the distribution arises from the graph structure via d-separation, not from canceling parameter values. This rules out "knife-edge" cases where causal effects precisely cancel. For our later results, these conditions need not be assumed; purely probabilistic analogues can emerge from an agent's credences.

[14]Mutual independence is conventional. Models with correlated exogenous variables can always be reformulated with independent ones by introducing additional variables.

[15]As we discuss in §7, our results don't rely on assumptions, like the causal Markov condition or causal faithfulness, that reference causal structure.

The irreducibility claim seeks to establish the Pearl Causal Hierarchy (PCH), introduced by Pearl and Mackenzie (2018), which distinguishes three levels of reasoning. Here we focus on the first two:

- **Associational Level** ($\mathcal{L}_1$): Encodes purely observational statements of the form $P(X = x \mid Y = y)$.
- **Interventional Level** ($\mathcal{L}_2$): Extends $\mathcal{L}_1$ with intervention statements like $P(X = x \mid \texttt{do}(Y = y))$, where the do-operator represents external manipulations.[16]

According to the PCH framework, the associational propositions expressed by the probability calculus ($\mathcal{L}_1$) are insufficient to capture the semantic content expressible by causal propositions of the do-calculus ($\mathcal{L}_2$). Pearl (2001) proffers:

> The bulk of human knowledge is organized around causal, not probabilistic relationships, and the grammar of probability calculus is insufficient for capturing those relationships. . . . (p. 31)

However, these claims of irreducibility remained informal until Bareinboim et al. (2022) formalized them by explicitly defining logical languages for each level of the hierarchy. They claim to vindicate Pearl's position by proving that these languages do not collapse into one another: higher-level languages are strictly more expressive than lower-level ones. In particular, they argue that interventional statements generally cannot be inferred solely from associational data:

> We prove that. . . it is generically impossible to draw higher-layer inferences using only lower-layer information, a result known informally in the field under the familiar adage: 'no causes-in, no causes-out.' (p. 513)

For Bareinboim et al., the stakes are clear:

> If each language did not expressively exceed its predecessors, then in some sense our talk of causation and imagination would be no more than mere figure of speech, being fully reducible to lower layers. (p. 527)

To evaluate expressivity rigorously, they treat $\mathcal{L}_1$- and $\mathcal{L}_2$-theories as *full deductive closures* of their atomic probability assignments in the respective languages.

The model M_1 from §3 illustrates the point. The chain $S \to G \to C$ entails the conditional independence $S \perp\!\!\!\perp C \mid G$, which determines an equivalence class of $\mathcal{L}_1$-theories: the set of $\mathcal{L}_1$ theories whose associations satisfy this independence. But a second model M_2 with the graph $S \leftarrow G \to C$—in which a genetic factor is the common cause of both smoking and cancer—entails the same conditional independence and can induce the same joint distribution: $P_{M_1}(S, G, C) = P_{M_2}(S, G, C)$. The two models are observationally indistinguishable. Yet they determine distinct $\mathcal{L}_2$-theories: they disagree on $P(C \mid \texttt{do}(S))$, since intervening on S changes C in M_1 but not in M_2, even though they agree on every $\mathcal{L}_1$-sentence.

[16] We set aside the counterfactual level ($\mathcal{L}_3$). Counterfactual reasoning in SCMs involves claims about what would have obtained under conditions that did not occur, and reducing such claims raises issues beyond those addressed by our enrichment strategy. Since $\mathcal{L}_2$ is sufficient for expressing causal decision theory (see §8), our main arguments about the role of causation in rational learning and action require only $\mathcal{L}_2$.

Bareinboim et al. show that this underdetermination of $\mathcal{L}_2$- by $\mathcal{L}_1$-theories is *generic*. To do so, they represent each structural causal model by a parameter vector in a finite-dimensional Euclidean space (encoding structural functions and exogenous distributions) and equip that space with standard Lebesgue measure. For any fixed observational theory $\mathsf{Th}_{\mathcal{L}_1}$, the class of models that realize it forms a measurable subset $\Omega_{\mathsf{Th}_{\mathcal{L}_1}}$. What drives their result is that interventional probabilities vary continuously with many structural parameters, whereas the induced $\mathcal{L}_1$-theory is locally stable under small changes that leave the observational distribution unchanged. Consequently, the models at which an $\mathcal{L}_1$-theory uniquely fixes its $\mathcal{L}_2$-theory form a null set. Thus observational equivalence but causal nonequivalence is overwhelmingly typical: generically, a single associational theory is compatible with many distinct interventional theories.

If this technical result forecloses reduction, then de Finetti's program fails for causation. We show that it does not. Our strategy is to drop the theorem's implicit assumption that languages are compared *over a fixed variable set*. We prove that a natural enrichment of the variables under consideration dissolves the apparent expressive gap.

5. Naturalizing Interventions

Having examined the case *against* reduction, we now demonstrate that reduction succeeds. We show how to extend any given causal model in a way that renders all interventional propositions expressible using probabilistic ones. This result enables a reduction of `do`-operator semantics to ordinary conditioning, without loss of expressive power or fidelity to the original model.

The key idea behind our first result—that exogenous interventions can be represented as endogenous variables in an expanded model—is not new. Pearl (2000) notes that a single `do`-operation in a model can be represented via conditionalization in an augmented model (§3.2 p. 71, equation 3.9). Dawid (2021, 2024) introduces regime indicator variables for interventions that express their content through conditional independence. And Spirtes et al. (1993) show that no two distinct directed acyclic graphs that are statistically indistinguishable remain so once suitably structured additional variables are made observable (§4, Theorem 4.6).[17] Our reduction theorem can be seen as a constructive refinement of these insights: for any SCM, we produce an explicit enriched model in which all of the empirical consequences encoded by interventional claims are recoverable by probabilistic ones.

What is novel is not this observation in isolation but rather its role in the larger argument. Combined with the representation theorem of §6 and the full algebra interpretation we develop below, it yields a thoroughgoing reduction of SCM-style causal reasoning to subjective probability.

Theorem 1 (Reduction Theorem)**.** *For any structural causal model, M, one can produce a probability function Q over a larger set of random variables that encodes all associational and interventional propositions true of M as conditional probability statements.*

[17] Two DAGs are *strongly statistically indistinguishable* if every distribution satisfying the causal Markov and minimality conditions on one satisfies those conditions on the other, and vice versa.

Proof. The strategy is as follows. We construct an extension M' of M such that all interventional propositions true of M are expressible as associational propositions in M', and all associational propositions of M remain true of M'. We then set Q equal to the joint distribution that M' induces over its variables. It remains to define M'.

Let $M = \langle U, V, F, P(U)\rangle$ be a structural causal model with exogenous variables $U = \{U_1, \ldots, U_n\}$, endogenous variables $V = \{V_1, \ldots, V_n\}$, structural equations $F = \{f_1, \ldots, f_n\}$ where $V_i = f_i(\mathrm{Pa}(V_i), U_i)$, and a joint distribution $P(U)$ over the exogenous variables. We produce an extension $M' = \langle U', V', F', P'(U')\rangle$ of M by introducing new exogenous intervention variables and modifying the structural equations so that conditioning on the intervention variables replicates the effects of `do`-operations. The procedure is as follows.

(1) **Intervention variables.** For each endogenous variable $V_i \in V$, introduce a new exogenous variable I_i. The range of I_i is $\mathrm{rng}(V_i) \cup \{\mathrm{NA}\}$, where the special value NA signifies that no intervention is performed on V_i.

(2) **Modified structural equations.** For each endogenous variable $V_i \in V$, replace the original structural equation $V_i = f_i(\mathrm{Pa}(V_i), U_i)$ with:

$$V_i = \begin{cases} I_i & \text{if } I_i \neq \mathrm{NA}, \\ f_i(\mathrm{Pa}(V_i), U_i) & \text{if } I_i = \mathrm{NA}. \end{cases}$$

When $I_i \neq \mathrm{NA}$, the variable V_i takes the value of I_i directly, overriding its original structural equation. When $I_i = \mathrm{NA}$, the original equation remains in force.

The extended model is $M' = \langle U', V', F', P'(U')\rangle$, where:

- $U' = U \cup I$, with $I = \{I_1, \ldots, I_n\}$. The exogenous variables of M' are the original exogenous variables together with the new intervention variables.
- $V' = V$. The endogenous variables are unchanged.
- $F' = \{f'_1, \ldots, f'_n\}$, the modified structural equations defined above.
- $P'(U') = P(U) \times \prod_{i=1}^{n} P(I_i)$, where each $P(I_i)$ is any distribution assigning positive probability to every element of $\mathrm{rng}(V_i) \cup \{\mathrm{NA}\}$.[18]

The product structure of $P'(U')$ ensures that the intervention variables are mutually independent and independent of the original exogenous variables U. In any SCM, the structural equations define a deterministic mapping from exogenous to endogenous variables; a distribution over the former therefore induces a unique joint distribution over all variables in the model (cf. Pearl 2000, p. 205). Accordingly, M' induces a joint distribution Q over $V' \cup U'$. We now verify the two required properties of Q.

Preservation of Associational Propositions. First, we show that all associational claims true of M are preserved in M'. Specifically, for any $X \subseteq V$ and $Y \subseteq V \cup U$:[19]

$$P(X = \bar{x} \mid Y = \bar{y}) \;=\; Q(X = \bar{x} \mid Y = \bar{y}, \mathbf{I} = \mathrm{NA}),$$

[18] The specific choice of $P(I_i)$ does not affect the reduction. All that is required is full support, so that conditioning on any value of I_i is well-defined. Indeed, since the perspective we develop shortly is one in which we start off with the full distribution Q, this non-uniqueness isn't an issue.

[19] Here and below, $\bar{x}$ and $\bar{y}$ denote vectors of realizations of the variables in X and Y, respectively.

where $\mathbf{I} = (I_1, \ldots, I_n)$. When $\mathbf{I} = \text{NA}$, every modified structural equation reduces to its original form: $V_i = f_i(\text{Pa}(V_i), U_i)$ for each i. That is, the enriched model M' behaves exactly as M does when no interventions are active. Consequently, the joint distribution on $V' \cup (U' \setminus I)$ conditional on $\mathbf{I} = \text{NA}$ in M' is identical to the unconditional joint distribution of $V \cup U$ in M, and the equality of conditional distributions follows.

Equivalence of Interventional Propositions. Next, we show that all interventional claims true of M are expressible as conditional probability statements in M'. Specifically, for any $X \subseteq V$ and $Y \subseteq V$:

$$P(X = \bar{x} \mid \texttt{do}(Y = \bar{y})) \;=\; Q(X = \bar{x} \mid \mathbf{I}_Y = \bar{y}, \mathbf{I}_{V \setminus Y} = \text{NA}),$$

where $\mathbf{I}_Y = \{I_i : V_i \in Y\}$ and $\mathbf{I}_{V \setminus Y} = \{I_i : V_i \notin Y\}$. Setting $\mathbf{I}_Y = \bar{y}$ forces each $V_j \in Y$ to take value $\bar{y}_j$, overriding its structural equation—this mirrors the graph surgery of $\texttt{do}(Y = \bar{y})$, which replaces the equations for Y with constants. Setting $\mathbf{I}_{V \setminus Y} = \text{NA}$ leaves the structural equations for all remaining endogenous variables in their original form. Since I is independent of U by the product structure of $P'(U')$, the distribution over U is unaffected by the conditioning on I. The enriched model therefore responds to $\mathbf{I}_Y = \bar{y}$ in the same way that M responds to $\texttt{do}(Y = \bar{y})$: the values of Y are fixed, their incoming equations are severed, and the rest of the model propagates accordingly. □

The theorem constructs a joint distribution Q over $V' \cup U'$ that encodes the full associational and interventional content of M. Crucially, both sides of the key equalities in the proof are ordinary conditional probabilities in Q. Associational propositions of M correspond to conditionals of the form $Q(\cdot \mid \cdot, \mathbf{I} = \text{NA})$, and interventional propositions correspond to conditionals of the form $Q(\cdot \mid \mathbf{I}_Y = \bar{y}, \mathbf{I}_{V \setminus Y} = \text{NA})$. In both cases, the right-hand side is an $\mathcal{L}_1$ statement about Q.[20] This yields the following.

Corollary 2. *For any structural causal model M, there exists a probability distribution Q over a set of random variables $\mathbb{X} \supseteq V' \cup U'$ such that every $\mathcal{L}_1$ and $\mathcal{L}_2$ proposition true of M is expressible as an $\mathcal{L}_1$ statement in Q.*

The $\mathcal{L}_2$ content of M is fully absorbed into the $\mathcal{L}_1$ content of Q.[21] An $\mathcal{L}_1$ theory describing associations over the enriched set of random variables is therefore enough to uniquely identify both the $\mathcal{L}_1$ and $\mathcal{L}_2$ statements true of an SCM defined over the strict subset of those variables encoded by V in M. Although M' still distinguishes exogenous from endogenous variables, this distinction plays no role once we pass to Q. The reduction target is Q, not M': in Q, every variable is simply a random variable, and no distinction between exogenous and endogenous remains.

[20] Strictly speaking, $\mathcal{L}_1$ theories are defined for SCMs, so speaking of the $\mathcal{L}_1$ theory of Q is a slight infelicity. We mean simply the set of associational statements true of Q as a probability distribution.

[21] This has the flavour of an expansion-conservativeness result: every model of the associational theory can be expanded to a model of the enriched theory. This connects our reduction to the literature on Ramsey sentences and conservative extensions in philosophy of science; see Button and Walsh (2018), ch. 3 for a systematic treatment. We leave exploration of this connection to future work.

Because interventions are now represented within the model rather than as external operations, we call them *naturalized interventions*. For each variable, one adds a naturalized intervention variable that, when inactive, retrieves the normal behavior of the original smaller model. When the intervention is active, it reproduces the effects of `do`-operations.

The construction above produces naturalized interventions that can be thought of as assignments of patients to the treatment or control arm of an experiment; or natural experiments where some population is exposed to a stimulus while another is not. In reality, naturalized interventions might be fine or coarse grained, they may be perfect, or may be partially confounded. To capture the idealized reasoning of the do-operator, we model interventions that are maximally fine grained and perfectly uncorrelated with other variables in the model. We state this more precisely.

Definition. *A variable I_X in a joint distribution Q over variables $\mathbb{X}$ is a* naturalized intervention *for $X \in \mathbb{X}$ relative to a fragment $\mathbb{F} \subseteq \mathbb{X} \setminus \{I_X\}$ if there exists a partition of $\mathbb{F} \setminus \{X\}$ into sets S and D such that: (1) $Q(X = x \mid I_X = x) = 1$ for all x in the range of X; (2) $I_X \perp\!\!\!\perp S$; (3) $D \perp\!\!\!\perp I_X \mid X$; (4) S is maximal with respect to condition (2).*[22]

All four conditions are stated entirely in the language of the probability calculus. (1) says that I and X are perfectly associated: knowing $I_X = x$ suffices to know $X = x$. (2) says that I_X carries no information about the variables in S. (3) says that, given X, knowing I_X provides no additional information about D—whatever association I_X has with D is fully accounted for by X. (4) ensures that S is as large as possible: it captures every variable in $\mathbb{F}$ that is unconditionally independent of I_X. In the graphical setting, S corresponds to the non-descendants of X and D to its descendants; but the definition makes no reference to graphs, directions, or causal ordering. The pair (I_X, S) is characterized simultaneously by the conditional independence structure of Q alone.

Despite differences in formulation, the three leading accounts of intervention in the causal modeling literature impose the same distributional constraints. That is, they agree on the empirical content of SCMs.

Woodward (2005, §3, p. 98) characterizes an intervention variable I on X with respect to Y by four conditions: (I1) I causes X; (I2) for some values, I screens off X from its other causes, acting as a switch for the value of X; (I3) every directed path from I to Y passes through X; and (I4) I is independent of every variable that causes Y along paths not through X. Pearl (2000, §3.2, pp. 90–92) enforces analogous constraints structurally: an intervention $\texttt{do}(X = x)$ replaces

[22] Without maximality, one could take $S = \emptyset$, and the conditions would be trivially satisfiable. In the models produced by Theorem 1, I_i is independent of all original exogenous variables and all other intervention variables. Since non-descendants of V_i are functions of subsets of U, they are jointly independent of I_X, so the maximal S exists and equals the non-descendants. A technical subtlety: joint independence of I_X from a set S does not follow from pairwise independence of I_X with each element of S, so in an arbitrary distribution Q, a unique maximal jointly independent set may not exist. To avoid this ambiguity, one can define S as the set of all variables in $\mathbb{F} \setminus \{X\}$ individually independent of I_X and then impose joint independence as a further condition. If joint independence fails, I_X does not qualify as a naturalized intervention. In the models produced by Theorem 1, the product structure of $P'(U')$ ensures that pairwise and joint independence coincide, so this complication does not arise.

the structural equation for X with the constant x and removes all incoming arrows into X, leaving all other structural equations and the distribution over exogenous variables intact. Spirtes, Glymour, and Scheines (1993, ch. 3, §§3.3–3.4, pp. 49–54) *represent* interventions by introducing *policy variables* that index experimental regimes. A policy variable must be exogenous with respect to the system variables, and changing its value constitutes a manipulation iff it alters their joint distribution; its children are the directly manipulated variables, whose parent sets—but no others—may change.

In each case, three distributional constraints emerge: the intervention fixes the value of its target, is independent of non-target causes, and affects other variables only through the target. Conditions (1)–(3) encode exactly these constraints as properties of a joint distribution. Condition (4) ensures that S reflects the full independence structure of Q rather than an arbitrary choice.[23]

Beyond these shared constraints, the three frameworks diverge, but none of their disagreements further constrains what interventions predict. Woodward builds modality into the definition of causation itself: for X to be a cause of Y there must there exist a possible intervention on X that would change Y when all other variables besides X and Y are held fixed. Causal claims thus presuppose the conceptual possibility of interventions. Pearl and Spirtes et al. impose no such condition. Pearl takes each structural equation to represent an autonomous mechanism, independently modifiable—a commitment that justifies graph surgery but adds nothing to the resulting distributions. Spirtes et al. connect the graph to probability via bridge principles like the Causal Markov and Faithfulness conditions that underwrite causal discovery from data but likewise leave predictions regarding interventions unchanged. On the distributional consequences of intervention, all three agree; as do we. Conditions (1)–(4) extract this common empirical core and discard the rest.[24]

From Enriched Algebras to Full Algebras. We proved Theorem 1 in one direction: given a structural causal model M, we constructed an enriched model M' that encodes all of M's interventional content as probabilistic conditioning. This construction served a specific purpose: it showed that the purported obstacles to reduction, formalized by Bareinboim et al. (2022), do not hold once the agent's representational resources are expanded. But the philosophical picture we wish to advance runs in the opposite direction.

Consider an agent who begins not with a causal model but with a probability distribution Q over some set of random variables $\mathbb{X} = \{X_i\}_{i \leq n}$. This is the agent's *full algebra*: the totality of random variables she takes to be relevant. In general this will be a very rich structure.

The full algebra is an idealization, but a familiar one. For example, Savage's (1954) framework describes a *grand world* decision problem since it posits a space rich enough to represent everything relevant to the agent's choice. Similarly, the

[23]Woodward's I4 is relativized to a particular outcome Y; condition (2) is stronger, requiring independence from all non-descendants. This matches Pearl and Spirtes et al., whose interventions are not outcome-relative. The product structure of $P'(U')$ in Theorem 1 yields the stronger condition.

[24]We focus on hard interventions. Soft interventions, which influence but do not fix variable values, can be accommodated by relaxing condition (1) while preserving (2) and (3); the reduction goes through mutatis mutandis.

Jeffrey-Bolker framework for decision theory requires a complete and atomless Boolean algebra over which an agent has preferences (1983). These are two examples of a familiar pattern: rationality principles are formulated at a level of richness that no actual agent attains, and we understand actual propositional attitudes such as belief and desire as approximating the ideal within a more limited representational scope.[25] Our full algebra is in the same spirit. At this level of description, Q is simply a joint probability distribution over random variables. Nothing about it is inherently causal.

Of course the agent need not, and typically cannot, reason about all of $\mathbb{X}$ simultaneously. Instead, she focuses on some strict subset of these variables, a *fragment* of Q. Interesting things can happen when one examines a fragment. Let $\mathbb{F} \subset \mathbb{X}$ be a fragment, and consider Q restricted to $\mathbb{F}$. There may exist variables in $\mathbb{X} \setminus \mathbb{F}$ that, viewed from within the fragment, satisfy the probabilistic conditions that characterize naturalized interventions: they covary perfectly with a variable in the fragment, they are independent of certain other variables in the fragment, and so on. When this obtains, the fragment admits a causal interpretation. The variables inside $\mathbb{F}$ can be organized into a structural causal model, and the variables outside $\mathbb{F}$ that satisfy the intervention conditions play the role of `do`-operations on that model.[26]

This is the perspective we endorse. Causal structure is not a feature of the world that our probability distributions must accommodate.[27] It is a feature that *emerges* when an agent, reasoning over a rich probability space, restricts attention to a fragment of it. But Q is the kind of distribution that, when one draws a boundary around some variables and calls them "in," can yield patterns that look interventional and hence causal. It is the conditional independence structure of Q that determines what kind of causal interpretation, if any, a given fragment supports.

Perhaps surprisingly, Pearl (2000) himself articulates a view along these lines:

> If you wish to include the entire universe in the model, causality disappears because interventions disappear—the manipulator and the manipulated lose their distinction. However, scientists rarely

[25] Icard (2023) describes this in terms of *levels of abstraction* (section 7.2). Huttegger (2017) discusses the connection between fully Bayesian models of learning and decision making over rich spaces and the methods that more bounded agents might deploy to approximate Bayesian reasoning.

[26] We call such a fragment a *causal fragment* of Q. These conditions may be satisfied only partially: some variables in the fragment may have corresponding naturalized interventions in $\mathbb{X}\setminus\mathbb{F}$ while others do not, and the intervention conditions themselves may hold only approximately rather than exactly. A systematic investigation of these intermediate cases is a natural direction for future work.

[27] Our view differs from others in the literature. Zhang et al. (2021) consider subjectivizing causal networks as well, and raise the question of what it is for an agent to *believe* a causal structure. They suggest that the answer should be given in terms of suppositional credences, drawing on the agency theory of causation developed by Price (1991) and Menzies and Price (1993). On their approach, believing a causal structure amounts to having a certain pattern of credences conditional on hypothetical interventions, where these interventions are understood as contemplated acts. Our framework provides an alternative: on the de Finetti line we pursue, believing a causal structure amounts to having exchangeable credences whose mixture representation concentrates on distributions exhibiting the corresponding conditional independence patterns. No primitive notion of agency or intervention is required.

> consider the entirety of the universe as an object of investigation. In most cases the scientist carves a piece from the universe and proclaims that piece in—namely, the focus of investigation. The rest of the universe is then considered out or background. . . . This choice of ins and outs creates asymmetry in the way we look at things, and it is this asymmetry that permits us to talk about "outside intervention" and hence about causality and cause-effect directionality. (p. 419)

Our results give this observation a precise formulation. A model containing the whole universe matches our full algebra perspective. The choice of ins and outs is the choice of a fragment. Naturalized interventions are variables left out. Causal structure is the conditional independence pattern that the fragment inherits from the full distribution. Causal structure reduces to probabilistic structure and vanishes as the model is sufficiently enlarged.

This also clarifies the role of Theorem 1. The theorem starts from a causal model and expands it, but this is best understood as a ground-clearing exercise. Its purpose is to demonstrate, against the backdrop of the irreducibility results, that no expressive power is lost in the passage from causal to probabilistic vocabulary. The constructive direction of the proof should not obscure the reductive direction of the conclusion: the ideal agent reasons in the full algebra, and causal models are useful approximations of what happens there when she instead must focus on a fragment of it.

Of course, agents are bounded. No one reasons over the full algebra of all variables they take to be relevant, just as no actual agent has perfectly exchangeable credences over an infinite set of random variables or satisfies the axioms of expected utility theory. The full algebra is an idealization of the same kind that pervades foundational work in formal epistemology and decision theory. Our claim is that at this level of idealization, causal structure drops out. Causal models pay rent by helping to manage boundedness: they allow an agent to reason about the effects of interventions on a fragment without tracking every variable outside it. The utility of causal reasoning reflects the limitations of the reasoner, not the structure of the world. We return to this theme in §8, where we consider how the choice between evidential and causal decision theories looks from the perspective of the full algebra.

6. Causal Reasoning Representation Theorem

The full algebra perspective raises a natural question about the status of the probability distribution Q itself. When the conditional independence structure of Q makes a fragment look causal, these patterns may appear objective, as if they reflect genuine structure in the world. Some might stop here, interpreting our result as reducing causation to a chance distribution over a larger set of random variables. We go a step further. Following de Finetti, we show that this distribution can itself be understood as emerging from an agent's exchangeable credences, completing the reduction from causation to *subjective* probability.

To do this we first state and prove a lemma, and then the core theorem. In conjunction with theorem 1 they show how to reduce much of causal reasoning to Bayesian reasoning.

Lemma 3 (Causal Structure Partition)**.** *Let $(X_1, X_2, \ldots, X_n)$ be n random variables, each taking values in a finite set. The simplex of all possible joint probability distributions over these variables can be partitioned into cells $\{R^k\}_{k=1}^K$ such that all distributions within each set R^k induce identical conditional independence relations among the variables.*

Proof. Let Δ be the simplex of joint distributions over the finite product space of $X_1, \ldots, X_n$. Define $\theta \sim \theta'$ iff θ and θ' induce identical conditional independence relations among the variables. This is an equivalence relation, and its equivalence classes yield the desired partition. □

The term *causal structure* in the name of this lemma connects to the perspective developed in § 5. Each cell R^k of the simplex corresponds to a pattern of conditional independence that, when restricted to a fragment, may admit a causal interpretation. The lemma partitions the space of possible distributions by these patterns, allowing the representation theorem to treat learning about causal structure as learning which cell one is in.

We can now state and prove:

Theorem 4 (Representation Theorem for Causal Inference)**.** *Let P be an agent's probability function over an infinite sequence of observations $D = \{D_1, D_2, \ldots\}$, where each $D_t = (X_1^t, X_2^t, \ldots, X_n^t)$ is a joint observation of all n random variables at time t, and each variable X_i takes values in a finite set. If P satisfies exchangeability over time points, then:*

(1) There exists a probability measure μ over the simplex of joint distributions such that P can be represented as a mixture of i.i.d. sampling distributions.

(2) The agent's beliefs and inferences about both causal structure and conditional probability distributions correspond to Bayesian updating on μ given the observed data $D_1, \ldots, D_m$.

Proof. For part (1): By Lemma 1, we have a simplex Δ of all joint probability distributions over the product space Ω of our random variables.

By the assumption of exchangeability, for any finite subset $\{D_1, D_2, \ldots, D_m\}$ of observations and any permutation π of the indices, we have:

$$P(D_1, D_2, \ldots, D_m) = P(D_{\pi(1)}, D_{\pi(2)}, \ldots, D_{\pi(m)}) \tag{1}$$

The multinomial version of de Finetti's theorem then applies directly to this sequence, stating that there exists a probability measure μ over the simplex Δ such that:

$$P(D_1, D_2, \ldots, D_m) = \int_\Delta \prod_{k=1}^m P_\theta(D_k)\, d\mu(\theta) \tag{2}$$

for any finite m, where $P_\theta(D_k)$ is the probability of observing D_k under distribution θ.

For part (2): By Lemma 1, the simplex Δ is partitioned into cells $\{R^k\}_{k=1}^K$ where each cell contains joint distributions that induce identical conditional independence relations.

The measure μ assigns probability mass to each cell R^k, which represents the agent's belief in the corresponding pattern of conditional independencies. Since each pattern of conditional independencies corresponds to a Markov equivalence

class of causal structures, $\mu(R^k)$ represents the agent's degree of belief in that causal structure.

Within each cell R^k, the conditional measure $\mu(\cdot|\theta \in R^k)$ represents the agent's uncertainty about the specific parameters given a particular causal structure.

When the agent observes data $D_1, D_2, \ldots, D_m$, Bayesian updating on μ produces a posterior distribution:

$$\mu(\theta|D_1, D_2, \ldots, D_m) \propto \prod_{k=1}^{m} P(D_k|\theta) \times \mu(\theta) \tag{3}$$

This posterior simultaneously updates both:

(1) The probability of each cell R^k (belief about causal structure)
(2) The distribution over parameters within each cell (belief about specific conditional probabilities)

Therefore, the agent's beliefs about causal structure and conditional probability distributions correspond to standard Bayesian updating on μ given the observed data. □

The representation theorem reveals that causal learning is not fundamentally different from other forms of statistical learning. Just as de Finetti showed that beliefs about objective chances can be understood as emerging from patterns of exchangeability in subjective probabilities, our result shows that beliefs about causal structure similarly emerge from these same patterns. This strengthens our case for the reducibility of causal reasoning to probabilistic reasoning: not only can interventional claims be recast as probabilistic ones (as shown in Theorem 1), but the very process of causal learning itself can be understood as a special case of Bayesian learning under exchangeability.[28]

The Full Picture. Here is the full picture as it stands. An agent has a probability distribution Q over a rich set of random variables, her full algebra. By the representation theorem, if her credences over repeated observations satisfy exchangeability, they can be represented as a mixture over chance distributions, each encoding a definite pattern of conditional independence among the variables. By the reduction theorem, when the agent restricts attention to a fragment of her full algebra, and there exist variables outside the fragment satisfying the conditions for naturalized interventions, the conditional independence structure within each

[28]Lindley (1982) arrives at a similar perspective in the narrower context of randomized experiments. After showing that Bayesian reasoning about experimental results can be grounded entirely in judgments of exchangeability, he concludes:

> No notion of cause and effect has been introduced into our analysis. The notion that the treatment causes an increase in response is language that is ambiguous and yet totally avoided by the unambiguous judgment of exchangeability. Causation is therefore a personal matter, not an objective fact, and the recognition of this is an important aid in understanding the nature of the phenomenon we refer to as causation. (p. 444)

Lindley also emphasizes that when two agents make different exchangeability judgments, "only extended experimentation can bring people into agreement" (p. 444), anticipating the convergence story that our representation theorem makes possible for causal learning in general.

chance distribution can be expressed in causal terms. Causal structure over a fragment is thus recovered from probabilistic structure in the full algebra, mediated by the representation theorem. At no point does anything irreducibly causal enter.

One might object that something objective remains. Even granting that chance distributions are artifacts of the representation, the conditional independence relations within them are not arbitrary. Some variables really are screened off from others; some really do covary. Does this not amount to objective causal structure by another name? It does not. Conditional independence is a feature of the agent's uncertainty, not of the world. As the agent accumulates evidence, her posterior concentrates on a narrower region of the simplex, and the conditional independence structure shifts accordingly. What looked like a stable causal relationship under one body of evidence may dissolve under another. The patterns are real features of the agent's epistemic state, but they are not mind-independent features of reality.

The point is sharpest at the limit. Consider an agent with complete knowledge of the values of all variables in her algebra. Her distribution is degenerate: every variable is assigned probability 0 or 1. A degenerate distribution has no nontrivial conditional independence structure. Every variable is trivially independent of every other. There are no dependencies to screen off, no associations to break by conditioning, no interventional contrasts to draw. Just as chance vanishes for Laplace's demon, causal structure vanishes too. For an omniscient agent, there is nothing for an SCM to model. Causation, like chance, is a feature of incomplete information.

7. Kicking Away the Causal Ladder

Our reduction yields a joint probability distribution over random variables, interpretable via the representation theorem as an agent's degrees of belief over in-principle observables. Our claim is that no causal primitives remain: what structure can (optionally) be interpreted causally merely emerges from probabilistic content. Yet one might worry that this is not right and that we have merely relocated causal assumptions rather than eliminated them. Perhaps structural equations, intervention variables, the graphical vocabulary of parents and descendants, or an assumption of causal faithfulness smuggle causation back in. We address each objection in turn.

Several of the objections share the same structure: the objector begins by assuming that a causal structure exists and then observes that our construction makes use of concepts that, under that assumption, have causal content. But our reduction does not begin from the assumption of causal structure. Our reduction arrives at an austere ontology: observables encoded by random variables, and degrees of belief encoded by a joint distribution over them. Causation appears as emergent patterns of uncertainty—conditional independencies, screening-off relations, covariation under conditioning—all fully expressible in the language of the probability calculus. The causal vocabulary is a convenient gloss on these probabilistic patterns, not a prerequisite for them. This is the spirit of de Finetti's reduction of chance: exchangeability is a property of the agent's credences, expressible without reference to chances, and the representation theorem shows that an agent with exchangeable credences behaves *as if* she believes in chances. Our treatment of causal structure is analogous.

Indeed, from the full algebra perspective of §5, several of these objections lose much of their force. If the agent simply begins with a distribution over her full variable set, she never introduces structural equations or intervention variables in the first place. These are features of a causal model that one might use to describe a fragment of her distribution, not ingredients of the distribution itself. Nevertheless, we address each objection on its own terms.

Structural equations. One natural objection is that structural equations are themselves causal claims, and that our reduction does not eliminate them. Note, however, that no structural equations remain at the end of our reduction. The reduction of a structural causal model M to an enriched

model M' yields a probability distribution Q over the enriched set of random variables—a point in a simplex—that captures all of the empirical content of M' and can make all of the relevant distinctions regarding the original variable set V that the causal model M could. To be completely unambiguous: Q is the object that our reduction yields. M' is a stepping stone. Furthermore, from the full algebra perspective developed above, Q is simply the agent's distribution restricted to the relevant variables.

Consider the model M in which smoking causes cancer via genetic mutations: $S \to G \to C$. Applying our construction yields an enriched model M' with naturalized intervention variables I_S, I_G, I_C and augmented structural equations, which in turn encodes a joint distribution Q over all variables. In M, we distinguish $S \to G \to C$ from other members of its Markov equivalence class by the fact that an intervention on S changes C: $P(C) \neq P(C \mid \texttt{do}(S = s))$. In M', conditioning on I_S makes the same distinction: $P(C) \neq P(C \mid I_S = s)$. The bare distribution Q, which encodes the same information without the apparatus of structural equations, preserves the distinction: $Q(C) \neq Q(C \mid I_S = s)$.

The observable consequences of M, M', and Q are equivalent. But the $\mathcal{L}_1$-theory of Q jettisons structural equations, the distinction between exogenous and endogenous variables—indeed, everything but distributional facts about random variables. Structural equations are wholly absent once the reduction is complete.

Intervention variables. A second objection targets the intervention variables introduced in Theorem 1. One might contend that these variables are defined by their causal role—they intervene on other variables, bypassing natural mechanisms—and that introducing them therefore presupposes the very causal concepts we claim to eliminate.

The objection conflates the heuristic motivation for the construction with its formal content. We introduced intervention variables by analogy with experimental manipulations because this makes the construction intuitive. But the formal definition of a naturalized intervention variable I_X appeals only to probabilistic conditions: (i) $Q(X = x \mid I_X = x) = 1$ for all x in the range of X; (ii) $I_X \perp\!\!\!\perp S$; (iii) $D \perp\!\!\!\perp I_X \mid X$; and (iv) S is maximal with respect to (ii)—where S and D partition the remaining variables. These are properties of a joint probability distribution that require nothing beyond the $\mathcal{L}_1$-language of the probability calculus. No condition references structural equations, graph structure, or causal ordering; the pair (I_X, S) is characterized simultaneously by the conditional independence structure of Q alone.

The name "intervention variable" is suggestive, but the suggestion is dispensable. We could call I_X a *switching variable* or an *override variable* without altering any formal property. What matters is that there exists a random variable in the full space satisfying conditions (i)–(iv). Whether that variable corresponds to an experimenter's manipulation, a natural experiment, or a cosmic coincidence is irrelevant to the reduction. Its status is fixed entirely by its probabilistic relationships to other variables.

Faithfulness. Finally, one might object that our reduction requires a faithfulness assumption and that faithfulness is itself a causal condition. Recall that faithfulness requires every conditional independence in the probability distribution P of M is entailed by the graphical structure of M via d-separation. If our results hold only when faithfulness obtains, and faithfulness is a condition on how causal structure relates to probability, then we have not achieved a pragmatic reduction of causation.

Does our reduction require faithfulness? It does not. Neither theorem uses a faithfulness assumption. The reduction theorem (Theorem 1) is a constructive result: given any SCM, one can produce an enriched model in which interventional propositions are expressible as conditional probability statements. This construction works regardless of whether the original model satisfies faithfulness.

Faithfulness enters only when one wants to interpret the conditional independence relations as reflecting causal structure. On that picture, under faithfulness, conditional independence patterns in the data uniquely identify a causal structure (up to Markov equivalence); without it, some causal structures may be empirically indistinguishable.

Worrying about such cases misunderstands our result. If a causal structure has empirical consequences—if it constrains observable distributions, even in principle—then that empirical content is entirely captured by our reduction. Only structure that is entirely inert, in the sense that it could never manifest in any observable difference, resists reduction. But positing such empirically idle structure amounts to adding metaphysical surplus, and it is precisely this surplus that our framework declines to carry.

Taken together, these responses establish that our construction does not presuppose the causal concepts it aims to reduce. At every stage—structural equations, intervention variables, graphical vocabulary, faithfulness—what appears to be causal content can be restated as a property of a joint probability distribution over random variables. The causal ladder was useful in climbing to our results; having arrived, we can kick it away.

8. Causation and Agency

We have provided an epistemology for causal learning: an agent with exchangeable credences can be represented as learning about causal structure via ordinary Bayesian updating. But the reduction would be incomplete if causal structure remained indispensable to rational agency, since that would preserve a strong motivation for causal realism that our results leave untouched. Drawing on existing work in decision theory, we develop a picture on which causal reasoning plays the same role in agency that it plays in epistemology: it is a tool that helps bounded

agents approximate what they would conclude by reasoning over a richer probabilistic structure. This picture accommodates the intuitions supporting causal decision theory while locating their source in the boundedness of the reasoner rather than in the structure of the world. A full defense of this picture would require a more thorough treatment of specific decision problems than we can offer here, but we develop it in enough detail to show that it is coherent, that it parallels the epistemological reduction of earlier sections, and that it connects naturally to several existing approaches in the literature.

8.1. **The Evidentialist Baseline.** One natural option is to embrace the de Finetti spirit fully. On this view, chance, and, in our context, causation, is a convenient representation of certain symmetries in our degrees of belief, not an additional feature of the world. We ought not reify this artifact of a particular way of representing our uncertainty. Taking this perspective seriously provides support for evidential decision theory (EDT). If one of the random variables under consideration represents the agent's possible acts, she should maximize conditional expected utility by conditioning on this act. From this perspective, nothing essentially causal is happening. The agent may use different chance hypotheses in the simplex as aids in calculating what to do, but fundamentally she is just conditioning on one variable among others.[29]

Once we have endogenized interventions, and, more broadly, the decision-relevant variables, and are working in the full algebra, this perspective becomes particularly natural. The agent's act is represented by a variable in the model. Conditioning on that variable and maximizing expected utility is simply what it means to choose well given one's beliefs.[30] Causal vocabulary, on this view, is useful shorthand for

[29]The foregoing assumes that agents have credences over their own acts. This assumption is contested. Spohn (1977; 2007) and Levi (1997), among others, argue that rational agents should not assign probabilities to acts under their deliberative control. If one's act is subject to choice, they maintain, it should not simultaneously be an object of probabilistic belief.

We acknowledge this concern but do not attempt to resolve it here. Following Jeffrey (1983) and Bolker (1967), we think it reasonable to permit act probabilities. This fits naturally with our framework, in which acts are representable as random variables in the full algebra. Those who reject act probabilities will find the decision-theoretic applications of our reduction more limited. For those who accept them, our framework provides a unified treatment of causal learning and causal decision-making within a thoroughly Bayesian setting.

For further reading, see Joyce (2002) and Hájek (2016) who respond directly to Levi and Spohn's concerns about act probabilities. Herrmann (2023) gives an overview of the debate, and how it interacts with the kind of naturalizing project we pursue here.

[30]Lindley (1982) develops a related point about randomized experiments. He observes that the purpose of an experiment is to inform future action: a doctor must decide whether to give *you* the treatment, not just the trial participants. The connection between the experimental results and your case must be established somehow, and the mechanism by which you receive the treatment will differ from the mechanism in the trial. In a causal framework, one would need to argue that the relevant causal structure is preserved. But Lindley argues that randomization serves a different function: it creates conditions under which the agent can judge the experimental cases and future cases to be exchangeable. Once this judgment is in place, de Finetti's representation theorem connects past observations to future predictions without any causal vocabulary. On this view, RCTs are tools for grounding exchangeability judgments, and the inferential chain from experiment to action runs entirely through exchangeability. Our framework generalizes this picture: experimental results are one source of data that informs the agent's full distribution, and exchangeability, not causation, is what connects evidence to rational action (pp. 440–441).

reasoning about fragments of the full probability space, but adds nothing at the fundamental level.

We find this picture compelling, especially from a naturalistic perspective. Both epistemically and decision-theoretically the agent treats herself as part of the universe, just like anything else. Yet many decision theorists hold that causal structure should play a role in practical reasoning beyond what pure conditioning provides. Our reduction sheds light on these views as well. We turn to them now.

8.2. **Reframing Meek and Glymour.** Meek and Glymour (1994) argue that the dispute between evidential and causal decision theorists can be recast as a disagreement about the causal processes underlying choice rather than a disagreement about which decision rule to follow. On their view, if an agent represents herself as intervening on a causal system, one should run causal decision theory (CDT), and calculate expected value conditional on an agent intervening to perform an act. If she represents herself as not intervening, the agent should run EDT, and calculate expected value conditional on her act. The question becomes not which decision theory to adopt *in general*, but how agents should understand their relationship to the causal system at hand. The agent should then use the decision theory that fits her situation.[31]

Our reduction sharpens this picture. Once interventions are endogenized, the question "am I intervening?" becomes a question about the conditional independence structure of the agent's act variable relative to a privileged subset of random variables. To say an agent is intervening on a fragment is just to say that her act variable, viewed from within that fragment, satisfies certain conditional independence conditions: it is uncorrelated with the non-descendants of the variable it sets, it screens off that variable from other variables, and so on. These conditions are expressed entirely in the language of probability theory.

From the perspective of the full algebra, one simply runs evidential decision theory either way. Whether the act would "look like" an intervention on some smaller fragment is a fact about the conditional independence structure of the full distribution. CDT, on this view, is not a rival to EDT but a useful approximation of it when reasoning within a fragment of the agent's full algebra. When the agent restricts attention to a causal fragment and her act satisfies the conditional independence conditions of a naturalized intervention on that fragment, CDT-style reasoning within the fragment recovers what full-algebra EDT would have recommended: the `do`-operator reimports the independence structure that was lost when the agent narrowed her representational scope.

8.3. **Uncertainty and the Mixing Problem.** The Meek and Glymour picture raises a natural question: what should an agent do when she is uncertain whether she is intervening? Stern (2019) addresses this by extending their framework to handle such cases.[32] In our framework, this uncertainty becomes uncertainty

[31]Both Hitchcock (2016) and Stern (2019) endorse something broadly along these lines, although both are less sanguine about an irenic resolution.

[32]Stern argues that a flatfooted attempt to do this doesn't work, but that a more sophisticated version that builds on work in Stern (2017) does. Stern (2026) goes further, describing a way of rationalizing CDT recommendations that is consistent with maximizing conditional expected utility: the agent brackets possibilities where she does not represent herself as intervening. He argues this is sufficient for generating CDT recommendations, without defending it as the correct

about where one sits in the simplex of probability distributions. Some cells have the property that the agent's act variable satisfies the conditional independence conditions characteristic of a naturalized intervention relative to some fragment. Other cells do not. As Stern points out, one can then integrate over these possibilities with respect to the agent's credences, making Meek and Glymour's proposal work under uncertainty.

Here a substantive concern arises, which may have bothered readers earlier in the present article. The agent's credences, by the representation theorem, are a *mixture* over possible chance distributions. But conditional independence relations that hold within each component distribution need *not* hold in the mixture.[33] Causal structure, understood as patterns of conditional independence, washes out when one mixes across distributions.

This is not a new observation. Meek and Glymour (1994) noted that an agent who takes two causally unconnected coins to share the same unknown bias will have credences that violate the Markov condition with respect to the believed causal structure, since the mixture over bias hypotheses induces a dependence between the coin outcomes that does not exist within any single hypothesis. Zhang et al. (2021) develop this point systematically, arguing that it poses a serious challenge for any subjective interpretation of causal Bayesian networks.

For the evidential decision theorist who doesn't want to reify certain symmetries in her credences, this poses no problem. She simply conditions on her act and maximizes expected utility using her actual credences, mixing and all. But for those who think conditional independence structure should play a more prominent role in decision-making, mixing is a genuine difficulty. If causal structure is decision-theoretically relevant beyond informing unconditional credences, and if that structure disappears in the mixture, then using the mixture directly seems to lose something important.

There are at least two broad strategies for responding. The first, pursued by Zhang et al. (2021), is epistemic: replace the agent's precise credences with a *set* of probability distributions. If the agent's doxastic state is represented by the set of candidate chance distributions rather than their mixture, each member of the set can satisfy the Markov condition individually, even though no single mixture of them would. This preserves causal structure at the cost of precise probabilism.[34]

account of genuine choice. In our framework, this would amount to restricting integration to cells of the simplex where the act variable satisfies the conditional independence conditions for a naturalized intervention. We set aside this refinement here.

[33]See, for example, Seidenfeld and Wasserman (1993), who worry about how this interacts with imprecise credences.

[34]Stern (2026) argues that the imprecise credences approach faces difficulties: it can be permissible to treat two variables as correlated in one's credences while accepting a causal graph that implies their probabilistic independence. Instead, Stern preserves both precise credences and conditional weighting but strengthens the interventionist constraints on choice so that the independencies between the agent's act and its non-effects are preserved under linear mixtures. On this approach, the mixing problem is addressed not by changing the decision rule or the representation of belief, but by restricting which act variables count as choices. We set this option aside here, though it fits naturally within our framework: the strengthened constraints can be understood as conditions on the agent's act variable relative to the full algebra.

The second strategy is decision-theoretic: preserve precise credences but change how they are used in calculating expected utility. The representation theorem provides a partition of the simplex into specific chance distributions, each encoding a definite pattern of conditional independence. When calculating expected utility, one can integrate over this partition in two ways. One can use conditional weighting: $P(K|A)$, the probability of each chance distribution conditional on the act under evaluation. This yields pure evidential decision theory. The act affects credence about which distribution obtains, and conditional independence structure washes out in the resulting mixture. Alternatively, one can use unconditional weighting: $P(K)$, the unconditional probability of each chance distribution. The act does not affect the weighting.

The unconditional option handles the mixing problem at the level of decision theory rather than epistemology. The agent's decision-making procedure respects the conditional independence structure within each chance hypothesis by refusing to let her act shift the weights across hypotheses. Where Zhang et al. preserve the Markov condition by enriching the representation of belief, this approach preserves precise belief by enriching the decision rule.

This is precisely the approach to causal decision theory developed by Skyrms (1984).[35] Skyrms takes expected utility to be:

$$U(A) = \sum_K P(K) \sum_C P(C \mid A, K) V(C)$$

where K ranges over cells of a partition representing possible objective chance distributions, C specifies the relevant consequence, and V is the value of the consequence for the agent. The key feature: $P(K)$ is unconditional on A. Within each cell, one conditions on the act. But the weighting across cells does not depend on which act is being evaluated.

Skyrms derives his K partition via de Finetti's representation theorem for chance (1984). Our representation theorem provides an exactly analogous partition for causal learning. Each K-cell corresponds to a specific joint distribution θ over the full variable set, encoding a definite pattern of conditional independence. An agent whose credences satisfy exchangeability is representable as uncertain over these distributions. The Skyrms formula then applies directly: integrate unconditionally over the chance hypotheses from our representation theorem, while conditioning on the act within each.

For those who find the mixing problem compelling, this combination of our reduction with Skyrms-style decision theory offers a response. One can accept that epistemically, one's credences are a mixture in which conditional independence structure does not hold. Yet one can preserve that structure for decision-making by keeping the weights across chance hypotheses unconditional.

8.4. **Our Considered View.** The preceding subsections traced how several existing approaches to decision theory look from within our framework. Here we draw these threads together. Our reduction was designed to undermine one prominent motivation for causal realism: the apparent indispensability of causal structure to prediction and learning. We showed that causal reasoning *is* dispensable once the agent's representational resources are enriched. A parallel motivation runs through

[35] Stern (2019) also takes an unconditional expectation.

agency: it seems that causal structure is indispensable to rational decision-making, since CDT appears to require it and the intuitions supporting CDT in cases like Newcomb's problem are strong. Our framework yields a natural account of the role of causation in decision-making, one that completes the picture developed in earlier sections.

We think our framework undermines this motivation in much the same way. Consider an agent reasoning in the full algebra of § 5. Her act is one random variable among others. She conditions on it and maximizes expected utility. This is EDT applied to a sufficiently rich probability space, and no causal vocabulary enters. The question is why, if this is all there is to it, causal reasoning seems so important in practice.

Our answer is that no agent reasons over the full algebra. She works within a fragment: a causal model covering the variables she takes to be most relevant. In narrowing her scope, she loses conditional independence structure that was present in the larger space. In particular, she may lose the information that would have told her how her act relates to other variables in the fragment. The `do`-operator compensates for this loss. When her act is well modeled as an intervention on the fragment, CDT reimports the independence structure that fragmentation discarded, and its recommendations approximate what full-algebra EDT would have delivered.

The Stavros example from the introduction illustrates the pattern. Stavros works in a fragment containing only "steak on menu" and "steak delivered." Within this fragment, his act of writing on the menu might seem like an intervention on the first variable. But in the full algebra, his act and the restaurant's menu listing are distinguished: they have different conditional independence properties. Had Stavros reasoned over the richer space, ordinary conditioning would have revealed his mistake. His error was not a failure of associational reasoning but a failure of scope.

This is the Meek and Glymour insight given precise content by our reduction. Whether an agent is "intervening" is a claim about whether her act variable satisfies certain conditional independence conditions relative to the fragment. When those conditions hold, CDT is a reliable approximation. When they do not, the approximation breaks down. Our framework also clarifies why approaches like the Skyrms K-partition are attractive: for those who take the chance distributions from the representation theorem seriously as constraints on decision-making, unconditional weighting across chance hypotheses offers a natural way to preserve conditional independence structure that mixing would otherwise wash out. But on our view, this reification is unnecessary. The representation theorem shows that an agent with exchangeable credences can be *represented as if* she believes in these chance distributions, not that the distributions themselves should play a role in her decision-making beyond generating her credences. Just as de Finetti held that chances are artifacts of the representation theorem rather than objects to be deferred to, we hold the same about the chance distributions over causal structures. The agent's credences are what they are. She should condition on her act and maximize expected utility.

The parallel to our broader argument then is this. Just as the apparent indispensability of causal structure to learning dissolves once one sees it as emerging from a richer probability space, the apparent indispensability of causal reasoning

to decision-making dissolves once one sees CDT as approximating EDT over a larger fragment. Our framework provides an account of the role of causation in decision-making on which that role is heuristic rather than foundational: causal reasoning helps bounded agents approximate what they would conclude by reasoning over a richer space. This does not settle the debate between EDT and CDT. But it does show that the intuitions supporting the importance of causal reasoning in agency are compatible with the thoroughly probabilistic picture we have been advancing, and that this picture offers a natural account of why causal reasoning is so useful in practice.

9. Conclusion

Attempts to reduce causation to patterns among observables have a long history. Hume argued that experience reveals only constant conjunction, not any further binding force (1739, §VII). Reichenbach pursued this line in probabilistic terms, proposing probability-raising as the core empirical signature of causation (1956).[36] De Finetti went further, envisioning a thoroughgoing subjectivism in which causal reasoning, like reasoning about chance, reduces to patterns in an agent's degrees of belief. We have completed this project for structural causal models.

Our reduction theorem shows that every causal model employing the `do`-calculus can be recast using only probabilistic conditioning once the agent's variable set is sufficiently expanded. Our representation theorem shows that an agent with exchangeable credences can be represented as learning causal structure via standard Bayesian updating. Just as de Finetti showed that chance-talk can be understood as emerging from exchangeable credences rather than worldly propensities, causal reasoning in SCMs emerges from probabilistic structure over the agent's full algebra.

As with de Finetti's reduction of chance, our results do not force eliminativism about causal relations. One can accept the formal results while maintaining that causation is a genuine feature of the world. But one prominent reason for positing causal structure as ontologically primitive has been that probabilistic reasoning alone seemed unable to capture it. Our results remove this motivation. Moreover, our reduction is designed to recover everything empirically relevant and pragmatically useful about causal reasoning within SCMs. Whatever causal structure a realist might still wish to posit beyond what our framework captures would be empirically idle. This does not refute causal realism, but it does make the case for it harder to sustain.

None of this is to say that causal reasoning is dispensable in practice. Causal reasoning remains pervasive in human cognition (Danks 2014), and for good reason: it allows agents to draw useful conclusions about a complex probabilistic landscape without having to represent that landscape in full. Our framework does not diminish this role but helps clarify it. Causal models, on our view, are tools that let bounded agents simulate reasoning over a larger space by reasoning over a smaller one. Understanding them this way is compatible with taking them to be among our most valuable cognitive tools, while recognizing that their value is practical rather than metaphysical.

[36] See also Russell (1913) and Suppes (1970).

This reframing also bears on recent debates about whether systems capable of associational reasoning, such as large language models, can reason causally (Zečević et al. 2023). If causal reasoning is a heuristic that helps bounded agents approximate what they would conclude by reasoning over a richer probabilistic structure, then the question is not whether a system implements the `do`-calculus but whether it arrives at the conclusions causal reasoning was helping us reach, perhaps by different means. A system that lacks explicit causal machinery but reasons well over a sufficiently rich set of associations may already be arriving at the conclusions that causal reasoning is designed to secure.

Within SCMs, causal talk is not fundamental. Whatever else causation may be, in this context it reduces to probability.

References

Arntzenius, F. and N. Hall (2003). On what we know about chance. *The British Journal for the Philosophy of Science 54*(2), 171–179.

Bareinboim, E., J. D. Correa, D. Ibeling, and T. Icard (2022). *On Pearl's Hierarchy and the Foundations of Causal Inference* (1 ed.)., pp. 507–556. New York, NY, USA: Association for Computing Machinery.

Bolker, E. D. (1967). A simultaneous axiomatization of utility and subjective probability. *Philosophy of Science 34*(4), 333–340.

Briggs, R. (2009). The anatomy of the big bad bug. *Noûs 43*(3), 428–449.

Briggs, R. (2012). Interventionist counterfactuals. *Philosophical studies 160*(1), 139–166.

Button, T. and S. Walsh (2018). Ramsey sentences and Newman's objection. In *Philosophy and Model Theory*, Chapter 3, pp. 55–73. Oxford: Oxford University Press.

Danks, D. (2014). *Unifying the Mind: Cognitive Representations as Graphical Models*. The MIT Press. MIT Press.

Dawid, A. P. (2021). Decision-theoretic foundations for statistical causality. *Journal of Causal Inference 9*(1), 39–77.

Dawid, P. (2024). What is a causal graph? *Algorithms 17*(3), 93.

de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré 7*, 1–68.

de Finetti, B. (1964). Foresight: Its logical laws, its subjective sources. In H. E. Kyburg and H. E. Smokler (Eds.), *Studies in Subjective Probability*, pp. 53–118. New York: John Wiley & Sons. Translated from the Italian (1937).

Eagle, A. (2025). Supplement A. Some Basic Principles About Chance. In E. N. Zalta and U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy* (Fall 2025 ed.). Metaphysics Research Lab, Stanford University.

Fishburn, P. C. (1981). Subjective expected utility: A review of normative theories. *Theory and decision 13*(2), 139–199.

Galles, D. and J. Pearl (1998). An axiomatic characterization of causal counterfactuals. *Foundations of Science 3*(1), 151–182.

Gallow, J. D. (2024). Decision and foreknowledge. *Noûs 58*(1), 77–105.

Hájek, A. (2016). Deliberation welcomes prediction. *Episteme 13*(4), 507–528.

Halpern, J. Y. (2013). From causal models to counterfactual structures. *Review of Symbolic Logic 6*(2), 305–322.

Herrmann, D. A. (2023). *Naturalizing Decision Theory*. University of California, Irvine.

Hertz, H. (1899). *The Principles of Mechanics Presented in a New Form*. Macmillian and Company, Limited.

Hitchcock, C. (2016). Conditioning, intervening, and decision. *Synthese 193*(4), 1157–1176.

Hume, D. (1739). *A treatise of human nature*. Oxford University Press.

Huttegger, S. M. (2017). *The probabilistic foundations of rational learning*. Cambridge University Press.

Icard, T. F. (2023). Resource rationality.

Jeffrey, R. C. (1983). *The logic of decision*. University of Chicago press.

Joyce, J. M. (1998). A nonpragmatic vindication of probabilism. *Philosophy of science 65*(4), 575–603.

Joyce, J. M. (1999). *The foundations of causal decision theory*. Cambridge University Press.

Joyce, J. M. (2002). Levi on causal decision theory and the possibility of predicting one's own actions. *Philosophical Studies 110*(1), 69–102.

Kyburg, H. E. (1978). Propensities and probabilities. *Dispositions*, 277–301.

Levi, I. (1997). *The covenant of reason: rationality and the commitments of thought*. Cambridge University Press.

Levinstein, B. A. and J. Spencer (2025). Bigger, badder bugs. *Mind 134*(533), 134–170.

Lewis, D. (1980). A subjectivist's guide to objective chance. In *IFS: Conditionals, belief, decision, chance and time*, pp. 267–297. Springer.

Lindley, D. V. (1982). The role of randomization in inference. In *PSA: Proceedings of the Biennial meeting of the philosophy of science association*, Volume 1982, pp. 431–446. Cambridge University Press.

Meek, C. and C. Glymour (1994). Conditioning and intervening. *The British journal for the philosophy of science 45*(4), 1001–1021.

Menzies, P. and H. Price (1993). Causation as a secondary quality. *The British Journal for the Philosophy of Science 44*(2), 187–203.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Pearl, J. (2001). *Bayesianism and Causality, or, Why I am Only a Half-Bayesian*, pp. 19–36. Dordrecht: Springer Netherlands.

Pearl, J. and D. Mackenzie (2018). *The Book of Why: The New Science of Cause and Effect* (1st ed.). USA: Basic Books, Inc.

Pettigrew, R. (2016). *Accuracy and the Laws of Credence*. Oxford University Press.

Pettigrew, R. (2020). *Dutch book arguments*. Cambridge University Press.

Price, H. (1991). Agency and probabilistic causality. *The British Journal for the Philosophy of Science 42*(2), 157–176.

Ramsey, F. P. (1931). Truth and probability. In R. B. Braithwaite (Ed.), *The Foundations of Mathematics and Other Logical Essays*, pp. 156–198. London: Routledge and Kegan Paul.

Reichenbach, H. (1956). *The Direction of Time*. Berkeley, CA: University of California Press.

Russell, B. (1913). On the notion of cause. *Proceedings of the Aristotelian Society 13*, 1–26.

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

Seidenfeld, T. and L. Wasserman (1993). Dilation for sets of probabilities. *The Annals of Statistics 21*(3), 1139–1154.

Skyrms, B. (1984). *Pragmatics and Empiricism*. Yale University Press.

Spirtes, P., C. Glymour, S. N., and Richard (1993). *Causation, Prediction, and Search*. Mit Press: Cambridge.

Spohn, W. (1977). Where luce and krantz do really generalize savage's decision model. *Erkenntnis*, 113–134.

Spohn, W. (2007). The core of free will. *Thinking About Causes. From Greek Philosophy to Modern Physics*, 297–309.

Stern, R. (2017). Interventionist decision theory. *Synthese 194*(10), 4133–4153.

Stern, R. (2019). Decision and intervention. *Erkenntnis 84*(4), 783–804.

Stern, R. (2026). The chances of choices. *The British Journal for the Philosophy of Science*.

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

van Fraassen, B. C. (1977). Relative frequencies. *Synthese*, 133–166.

Woodward, J. (2005). *Making things happen: A theory of causal explanation*. Oxford university press.

Zečević, M., M. Willig, D. S. Dhami, and K. Kersting (2023). Causal parrots: Large language models may talk causality but are not causal. *Transactions on Machine Learning Research*.

Zhang, J., W.-Y. Lam, and R. De Clercq (2013). A peculiarity in pearl's logic of interventionist counterfactuals. *Journal of philosophical logic 42*(5), 783–794.

Zhang, J., T. Seidenfeld, and H. Liu (2021). Subjective causal networks and indeterminate suppositional credences. *Synthese 198*(Suppl 27), 6571–6597.