Universiteit
Leiden
The Netherlands

From inference to influence: applying causal game theory to complex security environments

Vonk, M.C.

Citation
Vonk, M. C. (2026, March 26). *From inference to influence: applying causal game theory to complex security environments*. Retrieved from https://hdl.handle.net/1887/4299782

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/4299782

Note: To cite this publication please use the final published version (if applicable).

From Inference to Influence: Applying Causal Game Theory to Complex Security Environments

Proefschrift

ter verkrijging van
de graad van doctor aan de Universiteit Leiden,
op gezag van rector magnificus prof.dr. S. de Rijcke,
volgens besluit van het college voor promoties
te verdedigen op donderdag 26 maart 2026
klokke 13.00 uur

door

Maarten Vonk

geboren te Rotterdam, Nederland

in 1993

Promotor:

Prof.dr. T.H.W. Bäck

Co-promotoren:

Dr. A.V. Kononova

Dr. T. Sweijs (Netherlands Defence Academy)

Promotiecommissie:

Prof. dr. K.J. Batenburg

Prof. dr. M.M. Bonsangue

Dr. A.W. Laarman

Prof. dr. A. Plaat

Dr. C. Doerr (Sorbonne University, France)

Prof. dr. M. Postma (Tilburg University, Netherlands)

Copyright © 2026 Maarten Vonk All rights reserved.

This dissertation was supported by The Hague Centre for Strategic Studies, which provided a research position that complemented my work as a guest PhD candidate at the Leiden Institute of Advanced Computer Science. Additional support was provided by the Ministry of Foreign Affairs and the Ministry of Defense of the Netherlands through the PROGRESS framework. Responsibility for the contents rests solely with the author and does not constitute, nor should it be construed as, an endorsement by the Netherlands Ministries of Foreign Affairs and Defense. The Ministries had no involvement in the design, implementation, or content of the work.

Contents

1 Introduction **1**

1.1 Research Questions 4
1.2 Contributions 5
1.3 Outline 6
1.4 Publications of the Thesis 7

2 Preliminaries **9**

2.1 Probability Theory 9
2.2 Graphical Models 11
2.3 Game Theory 13

3 Causality and Assumptions **15**

3.1 Introduction 16
3.2 Potential Outcome Framework 18
3.2.1 Potential Outcomes 18
3.2.2 Randomized Controlled Trials 19
3.2.3 Beyond Randomized Controlled Trials 21
3.2.4 Structural Causal Models 23
3.3 Associational Level 24
3.3.1 Bayesian Networks 24
3.4 Interventional Level 27
3.4.1 Causal Discovery 28
3.4.2 Identification and Inference 34
3.4.3 Discovery, Identification and Inference with More Relaxations . 39
3.4.4 Practical Guide to Causal Inference 41
3.5 Counterfactual Level 42

3.5.1 One-Step-Ahead Potential Outcomes 44
3.5.2 Counterfactual Models 46
3.5.3 Inference 47
3.6 Conclusion and Future Work 48

4 Causal Game Theory **49**
4.1 Introduction 50
4.2 Game Theory 51
4.2.1 Normal-Form Game 52
4.2.2 Extensive-Form Game 54
4.2.3 Bayesian Game 56
4.2.4 Practical Guide to Game Theory 57
4.3 Causal Game Theory 59
4.3.1 Influence Diagram 59
4.3.2 Multi-Agent Influence Diagram and Causal Game 60
4.3.3 Causal Bayesian Games 62
4.3.4 Practical Guide to Causal Game Theory 64
4.4 Conclusion and Future Work 65

5 Optimization of Causal Interventions **67**
5.1 Introduction 67
5.2 Methodology 70
5.2.1 Discretization and Parameter Learning Methods 72
5.2.2 BDD Encoding and Weighted Model Counting 73
5.2.3 Optimization 74
5.3 Experimental Setup 74
5.3.1 Evaluation Measures 75
5.3.2 Bayesian Network Description 76
5.4 Results 80
5.4.1 Scalability of Inference Method 80
5.4.2 Trade-off Computational Cost and Quality of Discretization and Inference 81
5.4.3 Optimization Performance 83
5.5 Conclusion and Future Work 85

6 Real-World Applications **87**
6.1 Environmental Conflict . . . 87
6.1.1 Hypotheses . . . 88
6.1.2 Data and Methods . . . 91
6.1.3 Results . . . 92
6.1.4 Conclusion and Future Work . . . 97
6.2 Hybrid Threats . . . 98
6.2.1 Methodology . . . 99
6.2.2 Experimental Design . . . 104
6.2.3 Results . . . 106
6.2.4 Conclusion and Future Work . . . 110

7 Conclusions and Future Work **111**
7.1 Summary . . . 111
7.2 Conclusions . . . 114
7.3 Future Work . . . 119

A Acronyms and Notation Conventions **123**
A.1 Acronyms . . . 123
A.2 Notation Conventions . . . 124

B Appendix Chapter 5 **126**
B.1 Heatmap Results . . . 126
B.2 Experimental Set-up . . . 128
B.3 Evaluation Measures . . . 130

C Appendix Chapter 6 **131**
C.1 Experimental Data of Deterring Hybrid Threat . . . 131

Bibliography **135**

Summary **155**

Samenvatting **157**

Acknowledgements **159**

About the Author **161**

Chapter 1

Introduction

Strategic studies has long grappled with the challenge of understanding complex interactions between actors operating in uncertain and contested environments. Scholars and practitioners examining military conflicts, diplomatic negotiations, or security decisions seek analytical frameworks to support policy interventions. The field now employs formal methodologies from economics, mathematics, and social sciences to analyze strategic behavior, recognizing that intuition alone cannot resolve strategic problems.

Game-theoretical approaches are well established within strategic studies, offering robust tools for analyzing strategic interaction and decision-making under adversarial conditions [144, 101, 18]. In contrast, applications of causal inference remain notably limited. This absence is striking, given that effective policy intervention relies on identifying the causal effects of actions and distinguishing them from spurious correlations. Without a clear understanding of underlying causal mechanisms, policy decisions risk being based on misleading evidence, potentially resulting in unintended or ineffective outcomes. Causal inference, therefore, provides an essential foundation for formulating interventions that are both evidence-based and contextually appropriate.

To understand why causal inference has been underutilized in strategic studies, it is essential to examine the methodological challenges that have historically made causal claims difficult to establish. The study of causality has evolved significantly over time, rooted in philosophical and empirical traditions. David Hume, in the 18th century, argued that causation could not be directly observed but only inferred through regular associations and temporal ordering. This foundational problem, the impossibility of simultaneously observing both what happens when an intervention occurs and what

would have happened without it, was later formalized as the *fundamental problem of causal inference*. Researchers have thus traditionally exercised caution in making causal claims. One major response to this problem was the development of *randomized controlled trials* (RCTs), which offer a systematic approach to identifying causal effects by randomly assigning units to treatment and control groups, thereby eliminating bias from confounding variables. However, RCTs are often impractical or unethical in many settings, limiting their applicability.

This limitation gave rise to a rich body of work focused on causal inference using observational data. Donald Rubin advanced the potential outcomes framework. This framework offers a formal structure for thinking about *counterfactuals*, hypothetical outcomes that would have occurred under different treatment conditions [202]. Judea Pearl developed causal graphical models and the do-calculus, which provided a systematic approach that employed visual representations and mathematical operations to identify causal relationships from observational data [178]. Pearl also made a crucial distinction between causal discovery and causal inference. While *causal discovery* is concerned with identifying the underlying causal structure between variables by examining their statistical relationships, *causal inference* is concerned with probabilistically estimating an outcome variable under possible alterations. Pearl further refined causal inference by classifying it into three levels of increasing complexity, as outlined in his *causal hierarchy* [27]. Addressing queries at the more advanced levels often requires specialized causal calculi, which transform complex queries into solvable components, a process known as *causal identification*. Causal identification enables the implementation of *causal interventions*, manipulations of one variable that induce changes in others. In parallel, Joshua Angrist and Guido Imbens developed techniques for leveraging natural experiments, particularly instrumental variables [117]. Instrumental variables are factors that affect treatment assignment but have no direct effect on the outcome, allowing researchers to infer causality. Their contributions were recognized with the 2021 Nobel Prize in Economics, underscoring the maturation of causal inference as a rigorous scientific field.

These theoretical advancements have found extensive application in medical research and epidemiology, where establishing causal relationships is fundamental to understanding disease mechanisms and evaluating treatment effectiveness. However, the social sciences have adopted these methodological innovations more slowly, despite their clear relevance for the analysis of complex social and political processes. In the social sciences, approaches to causality have been fundamentally shaped by interpretive traditions, most notably Max Weber's methodological framework. Weber distin-

guished between *verstehen* (interpretive understanding) and *erklären* (causal explanation), arguing that effective social scientific analysis required both approaches working in tandem rather than viewing them as opposing methodologies. He sought to develop causal explanations that incorporated actors' subjective orientations within their specific historical and cultural contexts. Weber's insights remain particularly relevant because they anticipate a fundamental problem that becomes especially pronounced in strategic studies: the need to account for how actors' subjective interpretations of causal relationships shape their strategic responses, which in turn alter the very causal dynamics being analyzed.

Building on these general difficulties in social science methodology, security studies, as an application of strategic studies, faces a core analytical challenge that makes causal models particularly problematic. Complex security environments feature the convergence of interrelated threats, including armed violence, political fragility, socioeconomic instability, environmental stress, and multi-domain conflict dynamics, all shaped by interactions among multiple state and non-state actors. These characteristics systematically violate the fundamental assumptions underlying standard causal inference methods. Most critically, such environments exhibit widespread interference effects, where policy interventions could spread beyond their intended targets, affecting neighboring actors and outcomes. This makes it impossible to isolate causal effects cleanly. Simultaneously, strategic interdependence means that actors continuously monitor and adapt their behaviors in response to observed interventions by others. When security policies are implemented, adversaries do not passively accept their effects but actively counter-adapt, developing new tactics, shifting resources, or altering their strategic calculations in ways that change the very causal relationships that analysts seek to understand. In essence, interference and strategic considerations, among other characteristics, cause standard causal models to fail in complex security environments.

This analytical challenge, combined with difficulties in both comprehension and implementation of sophisticated methods [41, 165, 99], has led to a persistent lack of causal inference applications in complex security environments. The mathematical complexity of existing approaches often renders them inaccessible to security scholars and practitioners, while their computational requirements can present additional barriers for policy-makers operating under time pressure and resource limitations. This implementation gap has prompted increasing calls to bridge the divide between theory and practice through structured dialogues and closer collaborations between practitioners and methodologists [223, 165].

Addressing this challenge is crucial for several reasons. First, the characteristics of complex security environments demand analytical approaches that can match this complexity with methodological sophistication. Without them, analysts risk providing misleading or incomplete policy advice that undermines effective intervention. Second, causal inference methods must demonstrate their adaptability beyond computer science domains to establish broader disciplinary relevance and prevent methodological isolation. Third, real-world applications in complex security contexts reveal methodological limitations and drive the development of more sophisticated causal models, advancing the field through exposure to practical demands.

To address this issue, this dissertation develops an integrated framework that enables the application of causal inference in complex security environments. It adapts core causal concepts to reflect the specific features of such contexts, introduces computationally efficient methods for estimating intervention effects, and derives insights from real-world empirical applications.

1.1 Research Questions

This investigation is structured around three main research questions, which are further broken down into the following sub-research questions.

RQ1: **What conceptual frameworks and assumptions are necessary to enable the modeling of causal game theory in complex security environments?**

- RQ1.1: What fundamental causal concepts are necessary for structuring and differentiating causal relationships, particularly in the context of Pearl's causal hierarchy?
- RQ1.2: What key assumptions underpin causal inference applications across Pearl's causal hierarchy?
- RQ1.3: What methods exist for integrating causal reasoning with strategic decision-making in complex security environments, and how can they be applied?

RQ2: **How can optimal causal interventions be computed with high accuracy while ensuring computational efficiency?**

- RQ2.1: How can inference be performed efficiently to accurately estimate the effects of causal interventions while maintaining computational feasibility?

RQ2.2: How can optimization techniques be integrated with causal inference to optimize over causal interventions efficiently under budget constraints?

RQ3: **How can the proposed (strategic) causal concepts be applied to complex security environments?**

1.2 Contributions

In addressing these research questions, the dissertation makes theoretical, methodological, empirical, and societal contributions. Theoretical and methodological contributions establish a framework for causal reasoning in strategic contexts, while empirical applications to hybrid threats and climate conflict constitute the societal contributions by providing actionable insights for policymakers and security practitioners.

This thesis makes two primary *theoretical* contributions. First, it systematically organizes fragmented causal inference methods within Pearl's causal hierarchy, which structures causal reasoning across three levels: association, intervention, and counterfactual. This mapping clarifies the assumptions required at each level and enables practitioners to align specific policy questions with appropriate causal tools. Second, the thesis integrates strategic interaction within causal concepts using probabilistic graphical models, synthesizing the connections between diverse game forms such as normal form games, extensive form games, and Bayesian games and their causal representations. This theoretical unification articulates the mathematical distinctions between these models, clarifying which causal and strategic structures are required for different analytical contexts.

Regarding the *methodological* contribution, the thesis develops novel computational methods that address implementation barriers preventing causal inference application in complex security contexts. This contribution consists of an efficient approach for approximating the effect of causal interventions in hybrid Bayesian networks, combining discretization with knowledge compilation techniques. The approximation method is embedded within an optimization framework that identifies the most effective interventions under resource constraints. These methodological innovations transform computationally intractable problems into practical policy analysis tools.

The thesis makes an *empirical* contribution through two applications in complex security environments. The Iraq environmental conflict analysis uses causal discovery methods to uncover causal mechanisms linking environmental variables to conflict. The analysis computes causal estimates while accounting for spatial interference be-

tween municipalities. The hybrid threat deterrence application shows how strategic causal models establish equilibria between deterring and attacking agents. Sensitivity analysis reveals which factors most influence equilibrium outcomes. These studies provide new empirical insights into the role of causal mechanisms, interference, and strategic interaction in complex security settings.

The thesis makes a *societal* contribution by providing policymakers with practical tools and evidence-based findings for complex security environments. The empirical applications offer findings for formulating interventions in environmental conflict situations, while accounting for causal mechanisms and regional spillover. In hybrid threat scenarios, the application show how policymakers can assess the likely effects of measures in the presence of adversarial responses. The thesis provides the analytical tools necessary to conduct such analysis in contested environments where traditional methods fail due to interference effects and strategic interdependence.

In summary, this dissertation makes the following contributions to both scientific knowledge and societal understanding of complex security contexts:

- **Insights**: Adaptation of causal inference concepts through the clarification of underlying assumptions and the integration of strategic elements. This provides the basis for the formalization of strategic causal concepts, which are essential for reasoning in complex security environments.
- **Techniques**: Development of computational methods for approximating optimal causal interventions in hybrid Bayesian networks through discretization and knowledge compilation. This enables applications of causal inference tools in resource-constrained security environments.
- **Applications**: Empirical demonstrations of causal frameworks in complex security contexts through environmental conflict analysis in Iraq, addressing interference, and hybrid threat deterrence modeling, incorporating strategic interaction. These cases demonstrate the policy relevance of the proposed methods.

1.3 Outline

The research questions are addressed across different chapters of this dissertation, which is structured as follows. Chapter 2 introduces key preliminaries in probability theory, graph theory, and game theory. Chapter 3 disentangles fragmented research on causality by mapping core concepts onto Pearl's causal hierarchy. This provides practitioners with a structured framework for selecting appropriate methods aligned with

the level of causal inquiry they seek to address, while making explicit the assumptions each level entails. In doing so, the chapter responds to RQ1.1 and RQ1.2. Chapter 4 addresses RQ1.3 by bringing together scattered research on (causal) game-theoretical concepts and providing a structured guide to their application for practitioners. Chapter 5 presents the methodological contribution of this research and addresses RQ2 by introducing a computationally efficient method for optimizing causal interventions. Specifically, it includes the introduction of an efficient causal inference method, supported by empirical validation of its accuracy and computational efficiency. Additionally, a technique for integrating this method into an optimization framework for multiple causal interventions is proposed. In Chapter 6, RQ3 is addressed through the application of the previously introduced concepts to case studies in environmental conflict and hybrid threats. Finally, Chapter 7 presents the concluding remarks and future research directions.

1.4 Publications of the Thesis

All but one of the studies presented in this thesis have been published in reputable peer-reviewed journals and conference proceedings, reflecting the scholarly contribution and relevance of the work. The remaining paper is currently under peer review. The contents of this thesis consist of the following papers.

[251] Maarten C Vonk, Ninoslav Malekovic, Thomas Bäck, and Anna V Kononova. Disentangling causality: assumptions in causal discovery and inference. *Artificial Intelligence Review*, 56(9):10613–10649, 2023.

[252] Maarten C Vonk, Mauricio Gonzalez Soto, and Anna V Kononova. Graphical models for decision-making: Integrating causality and game theory. *arXiv preprint arXiv:2504.13210*, 2025.

[249] Maarten C Vonk, Sebastiaan Brand, Ninoslav Malekovic, Thomas Bäck, Alfons Laarman, and Anna V Kononova. Balancing computational cost and accuracy in inference of continuous bayesian networks. In *International Conference on Probabilistic Graphical Models*, pages 361–381. PMLR, 2024.

[253] Maarten C Vonk, Diederick Vermetten, Jacob de Nobel, Sebastiaan Brand, Ninoslav Malekovic, Thomas Bäck, Alfons Laarman, and Anna V Kononova. Optimizing causal interventions in hybrid bayesian networks. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 245–256. Springer, 2024.

[250] Maarten C Vonk, Anna V Kononova, Thomas Bäck, and Tim Sweijs. Multi-agent influence diagrams to hybrid threat modeling. *The Journal of Defense Modeling and Simulation*, 0(0), 2025.

[154] Ninoslav Malekovic, Maarten C Vonk, Laura Birkman, Tim Sweijs, Anna V Kononova, and Thomas Bäck. Applying causality to environmental security in Iraq. *Scientific Reports*, 15(1):16198, 2025.

Chapter 2

Preliminaries

This chapter introduces preliminaries about probability theory, graph theory, and game theory. Frequently used acronyms and notation conventions are introduced in Appendix A.

2.1 Probability Theory

This section discusses preliminaries of probability theory and the notation conventions followed throughout the dissertation. Probability theory revolves around events to which probability can be ascribed.

Definition 2.1 (State Space)**.** The *state space* Ω is the set of possible outcomes of an event.

Definition 2.2 (Event)**.** An *event* is a subset of Ω to which probability can be ascribed. The collection of all events is denoted by $\mathcal{F}$.

Definition 2.3 (Probability Measure)**.** A *probability measure* is a function $P : \mathcal{F} \to [0, 1]$ that satisfies the following requirements:

- $P(\Omega) = 1$.
- For all countable disjoint $C_1, C_2, \ldots \in \mathcal{F}$, the following holds:

$$P\left(\bigcup_{i=1}^{\infty} C_i\right) = \sum_{i=1}^{\infty} P(C_i).$$

Definition 2.4 (Probability Space)**.** The triplet $(\Omega, \mathcal{F}, P)$ is called the *probability space*.

Definition 2.5 (Conditional Probability)**.** For $C_1, C_2 \in \mathcal{F}$ and $P(C_2) > 0$, the *conditional probability* of C_1 given C_2 is defined as:

$$P(C_1 \mid C_2) = \frac{P(C_1 \cap C_2)}{P(C_2)}.$$

Definition 2.6 (Chain Rule)**.** Let $C_1, C_2 \in \mathcal{F}$ with $P(C_2) > 0$. The *chain rule* of probability expresses the joint probability of two events as:

$$P(C_1 \cap C_2) = P(C_2)P(C_1 \mid C_2).$$

Definition 2.7 (Bayes' Rule)**.** For $C_1, C_2 \in \mathcal{F}$ and $P(C_2) > 0$, *Bayes' rule* relates conditional probabilities as:

$$P(C_1 \mid C_2) = \frac{P(C_2 \mid C_1)P(C_1)}{P(C_2)}.$$

The concepts of random variables and their probability distributions provide a structured way to model and quantify uncertainty in real-world phenomena.

Definition 2.8 (Random Variable)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space. A *random variable* X is a function $X : \Omega \to \mathbb{R}$ such that for every $B \subset \mathbb{R}$:[1]

$$X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\} \in \mathcal{F}.$$

Definition 2.9 (Probability Distribution)**.** Let $(\Omega, \mathcal{F}, P)$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be a random variable. The *probability distribution* of X is the probability measure P induced by X, which assigns probabilities to subsets of $\mathbb{R}$, and is defined as:

$$P(B) = P(X^{-1}(B)) = P(\{\omega \in \Omega \mid X(\omega) \in B\}), \quad \forall B \subset \mathbb{R}.$$

Random variables are denoted by capital letters and $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ denotes the set containing random variables X_i that take values x_i in the corresponding state space Ω_{X_i}. The probability that random variable X_i takes value x_i is denoted by

[1] Technically, this condition should hold for every B in the Borel σ-algebra on $\mathbb{R}$, not all subsets. A full treatment of Borel sets is beyond the scope of this dissertation.

$P(X_i = x_i)$ or, in short, $P(x_i)$. If X_i is discrete, the probability distribution is denoted by uppercase $P(X_i)$. When X_i is continuous, it is denoted by lowercase $p(X_i)$.

Definition 2.10 (Joint Distribution)**.** The *joint distribution* of random variables $X_1, \ldots, X_n$ describes the probability distribution of their simultaneous occurrences and is denoted by $P(X_1, \ldots, X_n)$ or $P(\boldsymbol{X})$.

Definition 2.11 (Marginal Distribution)**.** A *marginal distribution* of a single random variable X_i taking value x_i is obtained by *marginalizing* the other random variables $\boldsymbol{X}_{-i}$ out:

- $P(x_i) = \sum_{\boldsymbol{x}_{-i} \in \Omega_{\boldsymbol{X}_{-i}}} P(x_i, \boldsymbol{x}_{-i})$ when the random variables are discrete.
- $p(x_i) = \int_{\Omega_{\boldsymbol{X}_{-i}}} p(x_i, \boldsymbol{x}_{-i}) d\boldsymbol{x}_{-i}$ when the random variables are continuous.

Definition 2.12 (Expected Value)**.** The *expected value* of a discrete random variable X_i under probability distribution P is

$$\mathbb{E}[X_i] = \sum_{x_i \in \Omega_{X_i}} x_i P(x_i)$$

while the *expected value* of a continuous random variable X_i under probability distribution p is

$$\mathbb{E}[X_i] = \int_{\Omega_{X_i}} x_i p(x_i) dx_i.$$

2.2 Graphical Models

The following graph-theoretic definitions underpin the causal inference concepts used throughout this thesis.

A graph is denoted by $G = (\boldsymbol{V}, \boldsymbol{E})$, where $\boldsymbol{V} = \{V_1, \ldots, V_n\}$ is the set of vertices (or nodes) and $\boldsymbol{E} \subseteq \boldsymbol{V} \times \boldsymbol{V}$ is the set of edges.

Definition 2.13 (Directed, Undirected, and Partially Directed Graphs)**.** A graph $G = (\boldsymbol{V}, \boldsymbol{E})$ is called:

- *directed* if every edge $(V_i, V_j) \in \boldsymbol{E}$ has an assigned direction from V_i to V_j;
- *undirected* if no edge in $\boldsymbol{E}$ has a direction;
- *partially directed* if $\boldsymbol{E}$ contains both directed and undirected edges.

For directed graphs, a fundamental structural property is the presence or absence of cycles.

Definition 2.14 (Cycle)**.** A *cycle* in a directed graph is a sequence of vertices $V_{i_1}, \ldots, V_{i_k}$ where $k \geq 2$, $(V_{i_j}, V_{i_{j+1}}) \in \boldsymbol{E}$ for all $j = 1, \ldots, k-1$, and $(V_{i_k}, V_{i_1}) \in \boldsymbol{E}$.

Definition 2.15 (Directed Acyclic Graph)**.** A *directed acyclic graph* (DAG) is a directed graph that contains no cycles.

In addition to standard directed and undirected edges, bidirected edges are used to represent certain causal structures.

Definition 2.16 (Bidirected Edge)**.** A *bidirected edge* between vertices V_i and V_j is denoted by $V_i \leftrightarrow V_j$ and represents a specific edge type distinct from both directed edges $(V_i \to V_j)$ and undirected edges $(V_i - V_j)$.

Definition 2.17 (Acyclic Directed Mixed Graph)**.** An *acyclic directed mixed graph* (ADMG) is a graph $G = (\boldsymbol{V}, \boldsymbol{E})$ where $\boldsymbol{E}$ consists only of directed and bidirected edges, and the directed edges form no cycles.

The following relational concepts describe how vertices are connected within directed graphs.

Definition 2.18 (Parent and Child)**.** Let $G = (\boldsymbol{V}, \boldsymbol{E})$ be a directed graph. If $(V_1, V_2) \in \boldsymbol{E}$ (denoted $V_1 \to V_2$), then V_1 is called a *parent* of V_2 and V_2 is called a *child* of V_1. The set of parents of V_i is denoted by $\mathbf{pa}(V_i) = \{V_j \in \boldsymbol{V} \mid (V_j, V_i) \in \boldsymbol{E}\}$, and the set of children of V_i by $\mathbf{ch}(V_i) = \{V_j \in \boldsymbol{V} \mid (V_i, V_j) \in \boldsymbol{E}\}$.

Definition 2.19 (Ancestor and Descendant)**.** Let $G = (\boldsymbol{V}, \boldsymbol{E})$ be a directed graph. An *ancestor* of node V_i is any node V_j for which there exists a directed path from V_j to V_i, including V_i itself. The set of ancestors is denoted by $\mathbf{an}(V_i)$. Similarly, a *descendant* of V_i is any node V_j for which there exists a directed path from V_i to V_j, including V_i itself. The set of descendants is denoted by $\mathbf{de}(V_i)$.

Definition 2.20 (Non-descendants)**.** The set of *non-descendants* of V_i is defined as $\mathbf{nonde}(V_i) = \boldsymbol{V} \setminus \mathbf{de}(V_i)$. Note that $V_i \notin \mathbf{nonde}(V_i)$ since $V_i \in \mathbf{de}(V_i)$.

For acyclic graphs, a linear ordering can be imposed on the nodes that respects the edge directions.

Definition 2.21 (Topological Sort)**.** A *topological sort* is a total ordering $<$ on $\boldsymbol{V}$ such that $(V_i, V_j) \in \boldsymbol{E}$ implies $V_i < V_j$. A topological sort exists if and only if the graph is acyclic.

Two important graph operations involve restricting the vertex set (subgraphs) or removing specific edges (mutilated graphs).

Definition 2.22 (Subgraph)**.** Given a topological sort $<$ on $\boldsymbol{V}$, the *subgraph* G_i is the induced subgraph on the vertex set $\{V_j \in \boldsymbol{V} \mid V_j \leq V_i\}$, i.e., the graph restricted to nodes that precede or equal V_i in the topological ordering.

Definition 2.23 (Mutilated Graph)**.** Given a subset $\boldsymbol{V}' \subseteq \boldsymbol{V}$, the *mutilated graph* $G_{\overline{\boldsymbol{V}'}}$ is the graph $(\boldsymbol{V}, \boldsymbol{E}')$ where $\boldsymbol{E}' = \boldsymbol{E} \setminus \{(V_i, V_j) \in \boldsymbol{E} \mid V_j \in \boldsymbol{V}'\}$, i.e., all edges directed into nodes in $\boldsymbol{V}'$ are removed.

When graphs are endowed with probabilistic meaning, random variables $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ will correspond to nodes of the graph $\boldsymbol{V} = \{V_1, \ldots, V_n\}$ and therefore $\boldsymbol{V}$ will inherit the probability distributions and state spaces from $\boldsymbol{X}$ (meaning $P(\boldsymbol{V})$ and v_i will correspond to $P(\boldsymbol{X})$ and x_i, respectively). In this case, $\mathbf{pa}(V_i)$ refers to the random variables that are associated with the parents of V_i. The assignment of random variables $\mathbf{pa}(V_i)$ is denoted by $\boldsymbol{pa}_i$, which is an element of state space $\Omega_{\mathbf{pa}(V_i)}$.

2.3 Game Theory

A game models strategic interactions among rational decision-makers, called *agents* or *players*, where each decision-maker selects actions to maximize their own objectives while considering the choices of others. It provides a structured framework to analyze decision-making in competitive or cooperative settings.

Definition 2.24 (Game)**.** A *game* is denoted by $\Gamma = (M, \boldsymbol{A}, \boldsymbol{U})$ and consists of:

- A finite set of *players* $M = \{1, \ldots, m\}$.
- A set of *action sets* $\boldsymbol{A} = \{A^1, \ldots, A^m\}$, where A^i is the set of available actions for player $i \in M$.
- A set of *utility functions* $\boldsymbol{U} = \{u^1, \ldots, u^m\}$, where $u^i : \prod_{j \in M} A^j \to \mathbb{R}$ maps each action profile $(a^1, \ldots, a^m)$ with $a^j \in A^j$ to a real-valued payoff for player i.

An *action profile* is a tuple $a = (a^1, \ldots, a^m) \in \prod_{j \in M} A^j$ specifying one action for each player. For a given player i, the notation $a^{-i} = (a^1, \ldots, a^{i-1}, a^{i+1}, \ldots, a^m)$ denotes the *partial action profile* of all players except i, so that $a = (a^i, a^{-i})$.

Definition 2.25 (Strategy)**.** A *strategy* (or *mixed strategy*) for player $i \in M$ is a probability distribution $\sigma^i \in \Delta(A^i)$, where $\Delta(A^i)$ denotes the simplex over the action

set A^i. That is, σ^i assigns a probability $\sigma^i(a^i) \geq 0$ to each action $a^i \in A^i$ such that $\sum_{a^i \in A^i} \sigma^i(a^i) = 1$. The set of all strategies for player i is denoted by $\mathbf{\Sigma}^i = \Delta(A^i)$. A strategy is called *pure* if $\sigma^i(a^i) \in \{0, 1\}$ for all $a^i \in A^i$, meaning the player deterministically selects a single action, and *fully mixed* if $\sigma^i(a^i) > 0$ for all $a^i \in A^i$, meaning every action is chosen with positive probability.

A *strategy profile* $\sigma = (\sigma^1, \ldots, \sigma^m)$ specifies a strategy for each player. The expected utility for player i under a strategy profile σ is denoted $u^i(\sigma)$.

Definition 2.26 (Strategy Profile)**.** A *strategy profile* is a tuple $\sigma = (\sigma^1, \ldots, \sigma^m) \in \prod_{j \in M} \mathbf{\Sigma}^j$ specifying one strategy for each player. For a given player i, the notation $\sigma^{-i} = (\sigma^1, \ldots, \sigma^{i-1}, \sigma^{i+1}, \ldots, \sigma^m)$ denotes the *partial strategy profile* of all players except i, so that $\sigma = (\sigma^i, \sigma^{-i})$.

Chapter 3

Causality and Assumptions

To effectively apply causal research within complex security environments, it is essential to understand both the available causal concepts and the specific types of causal claims they support. However, causal inference research remains fragmented across multiple scientific disciplines, often leading to conceptual silos and inconsistent terminology. Synthesizing these disparate components is therefore necessary to build a coherent foundation for causal reasoning.

This chapter addresses that need by systematically organizing causal inference concepts within Pearl's causal hierarchy [27], which provides a formal structure for distinguishing three levels of causal reasoning: association, intervention, and counterfactual. Beyond mapping methods to this hierarchy, it also clarifies the assumptions required for rigorously applying causal models in practice and offers guidance for practitioners seeking to implement them effectively.

By adopting this structured approach, the chapter directly engages with the first research question, RQ1.1: *What fundamental causal concepts are necessary for structuring and differentiating causal relationships, particularly in the context of Pearl's causal hierarchy?* In addition, it explores RQ1.2: *What key assumptions underpin causal inference applications across Pearl's hierarchy?*

This dual focus equips practitioners with the foundational causal concepts needed to formulate meaningful and well-structured causal claims, while simultaneously cultivating a deeper understanding of the methodological commitments and assumptions required to substantiate them. The chapter's discussion closely follows a previously published article [251].

3.1 Introduction

Historically, the fundamental problem of causal inference, which arises from the impossibility of observing both the treated and control outcomes for the same unit, made it difficult for researchers to establish causal claims [102]. To overcome this challenge, randomized controlled trials (RCTs) became the gold standard for identifying causal effects, as the random assignment of units to treatment or control effectively eliminated confounding between assignment and outcome. However, in many contexts (e.g., studying the effects of smoking), it is either unethical or impractical to randomly assign individuals to treatment conditions. As a result, researchers must rely on *observational data* and alternative approaches to draw causal claims.

In this chapter, causal concepts that emerge from reasoning with causality using observational data within the context of *probabilistic graphical models* (PGMs) are explored. The latter are graphical representations that can be learned from observational data through causal discovery, an algorithmic approach to inferring the causal structure among variables. The next step, causal identification, determines whether a causal effect can be estimated from the available observational data, and if so, employs specific calculi to express causal queries in terms of known quantities, ensuring they have a unique solution [177]. With additional assumptions, one can perform causal inference, which involves estimating an outcome variable under hypothetical interventions. The focus is on the assumptions required at each stage, including causal discovery, identification, and inference, and on the different causal concepts that emerge from these processes.

Table 3.1: Pearl's Causal Hierarchy Queries

Level	Action	Query	Example
1. Associational $P(Y \mid x)$	Seeing	How does observing $X = x$ influence Y?	Do smokers generally tend to have more lung cancer than non-smokers?
2. Interventional $P(Y \mid do(x), z)$	Doing	How does intervening on $X = x$ affect Y given $Z = z$?	Is there a causal effect of smoking on lung cancer?
3. Counterfactual $P(Y(x) \mid x', y')$	Imagining	What would have been Y under $X = x$ given that $Y = y'$ is observed under $X = x'$?	Would a patient have lung cancer if he/she had smoked given that the patient does not have lung cancer and has never smoked?

Causal identification and causal inference can be further categorized into three levels of increasing complexity, known as Pearl's causal hierarchy [27]. These levels correspond to different types of queries: *associational* (seeing), *interventional* (doing), and *counterfactual* (imagining) [179]. Table 3.1 provides an overview of these queries, while Figure 3.1, adapted from Bareinboim et al. [27], presents a structured depiction of the key concepts at each level of Pearl's causal hierarchy, with higher levels shown towards the top of the figure to highlight increasing complexity. It has been proven by the *causal hierarchy theorem* [27] that queries at higher levels of the hierarchy can generally not be addressed with information of lower levels only.

To navigate this hierarchy, Section 3.2 introduces the *potential outcome framework* (POF), providing a useful perspective for analyzing key causal concepts and their foundational assumptions [202]. This is followed by a more formal, yet logically equivalent, approach that adopts a distinct notational and conceptual perspective: the *structural causal model* (SCM) [178]. Both serve as foundational frameworks for addressing causal queries across all three levels. Building on the SCM, its natural extension is introduced, the *spatially equivalent structural equation models* (SESEM), which allows for modeling in environments where interference may be present [137]. Then, Bayesian networks, d-separation, and some equivalent Markov assumptions at the associational level of the hierarchy are introduced in Section 3.3. Concepts and assumptions at the interventional level of the hierarchy will be introduced in Section 3.4. In Section 3.4.1, different sets of assumptions that allow non-parametric as well as parametric causal discovery are introduced. Subsequent Section 3.4.2 delineates different assumptions and concepts for non-parametric as well as parametric identi-

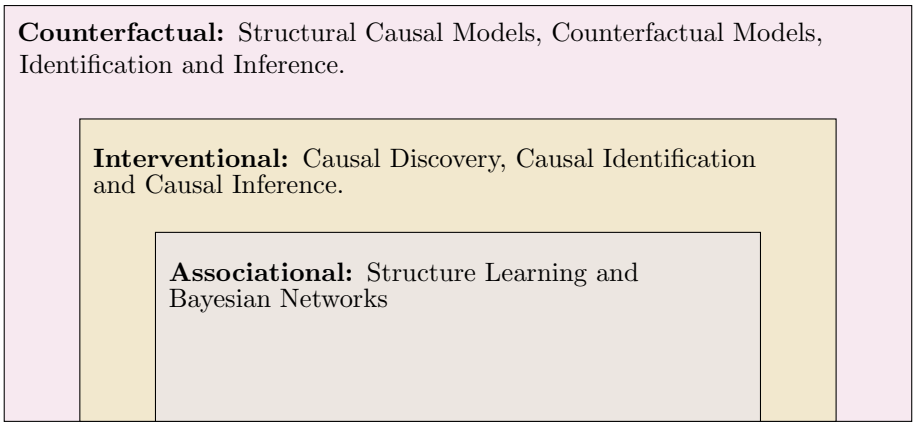


Figure 3.1: Pearl's causal hierarchy of causal concepts.

fication and inference approaches while enunciating the meeting point between the two approaches. The chapter proceeds with the introduction of various counterfactual models and inference techniques to enable reasoning at the counterfactual level of the hierarchy in Section 3.5. Finally, concluding remarks are presented in Section 3.6.

3.2 Potential Outcome Framework

In this section, the potential outcome framework (or Neyman-Rubin causal model) as developed by Rubin [202] is introduced. The potential outcomes ground the most granular sort of queries of the causal hierarchy, the counterfactual, and the framework incorporates the core assumptions of causal inference. That means that claims about potential outcomes are equivalent to counterfactual claims. The necessary methods and targets of interest will be defined along with accessory assumptions. For a full picture of these methods and assumptions, the reader is referred to Figure 3.2. This section naturally revolves around the concept of *potential outcomes*.

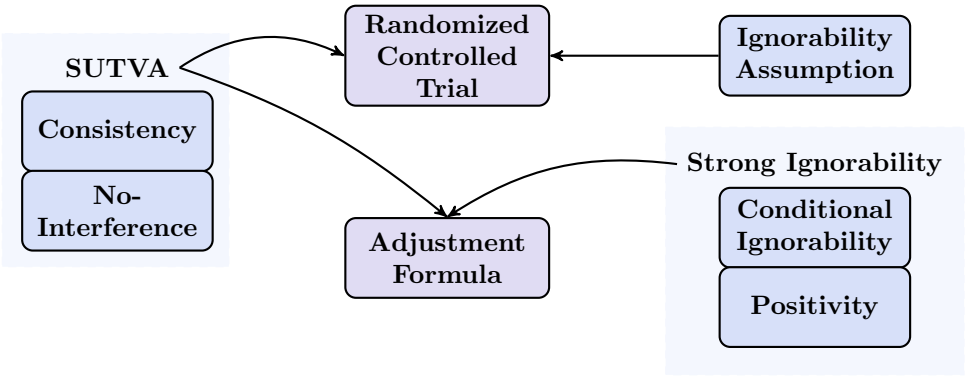


Figure 3.2: Methods for inferring causal claims under different assumptions. Boxes represent assumptions (individual or grouped); arrows indicate that satisfying the source enables the use of the target method. The ignorability assumption and stable unit-treatment value assumption (SUTVA) are implicit in randomized controlled trials for which causal claims can be made. When strong ignorability holds together with SUTVA, the adjustment formula should be invoked to calculate causal estimates. These concepts are further introduced in the following sections.

3.2.1 Potential Outcomes

Before the potential outcomes are introduced, first the treatment will be defined:

Definition 3.1 (Treatment Variable)**.** The treatment variable T is a random variable that takes on different values for treatment t.[1]

Definition 3.2 (Potential Outcome)**.** The potential outcome random variables are denoted by $Y(T = t)$ (or $Y(t)$ in short) for different treatment values $T = t$. For a unit of observation i (or unit in short) and treatment value t, the potential outcome realizations are denoted by y_i^t, the outcome that would have been observed if unit i had been exposed to treatment t.

Classically, t has been considered to take on binary values corresponding to treatment (1) and control (0) [202]. The first target of interest emerges naturally from this definition and is called the *unit-level causal effect*.

Definition 3.3 (Unit-Level Causal Effect)**.** Considering binary treatment t, the unit-level causal effect for unit i is defined as $\tau_i = y_i^1 - y_i^0$.

The potential outcome of unit i cannot be observed for treatment $t = 1$ and control $t = 0$ in a single observation, leading to the fundamental problem of causal inference [102]. This means that the unit-level causal effect cannot be calculated exactly but only estimated. y_i^t is called *counterfactual* when unit i has not been exposed to treatment t but to another treatment value $t' \neq t$. The unit-level causal effect also has its statistical population counterpart, the *average treatment effect*.

Definition 3.4 (Average Treatment Effect (ATE))**.** For binary treatment $t \in \{0, 1\}$, the average treatment effect is defined as

$$\tau = \mathbb{E}[Y(T = 1) - Y(T = 0)]. \tag{3.1}$$

For a sample of n units, the sample average treatment effect is $\hat{\tau} = \frac{1}{n} \sum_{i=1}^{n} (y_i^1 - y_i^0)$.

3.2.2 Randomized Controlled Trials

Randomized controlled trials are widely considered to be the gold standard for estimating average treatment effects, as they inherently satisfy three key assumptions. To state these assumptions, the observed outcome for unit i is introduced, denoted by y_i, which is the outcome actually measured after treatment assignment. Each unit reveals only one potential outcome; the one corresponding to the treatment actually received.

The first assumption is called *consistency*:

[1] Although the treatment variable can be multi-dimensional, the fundamental causal concepts are most clearly introduced using a single treatment variable. This concept is generalized in Section 3.5.

Assumption 1 (Consistency)**.** For each unit i that receives treatment t_i, the observed outcome equals the potential outcome under that treatment:

$$y_i = y_i^{t_i}.$$

Equivalently, at the random variable level: if $T = t$, then $Y = Y(t)$.

Informally, the assumption forces one to unambiguously define treatment and tie the potential outcomes to the observed variables. Although this assumption is sometimes known as the *no-multiple-treatment* assumption, some researchers draw a firm distinction between the two [245]. Consistency can be a strong assumption in the observational setting, but it is implicit in randomized controlled trials, because exposure to treatment is a result of experimental design [51].

The second assumption is known as the *no-interference* assumption [54]. It explicitly states that a potential outcome of a unit is not dependent on treatment received by other units. More formally,

Assumption 2 (No-Interference)**.** Let t_i be the treatment assignment of unit i for $i = 1, \ldots, n$. Then no-interference is satisfied if

$$Y_i(t_1, \ldots, t_n) = Y_i(t_i).$$

Interference is also known as *spillover*. In a randomized controlled trial, the investigator can prevent causal spillover by designing the experiment such that different units do not interact.

A combination of both consistency and no-interference leads to the *stable unit-treatment value assumption* (SUTVA) [203]. As interference is hard to restrain in the observational setting, a formal framework capable of addressing violations of interference assumptions will be introduced when presenting structural causal models in the next section. Although a randomized controlled trial poses limitations on SUTVA violations, the strength of the randomized controlled trial lies in its implication of the *ignorability* assumption:

Assumption 3 (Ignorability/Exchangeability)**.** Consider binary treatment assignment random variable T and potential outcome under treatment $Y(1)$ and control $Y(0)$. Then, ignorability is satisfied if

$$Y(0), Y(1) \perp\!\!\!\perp_P T,$$

where $\perp\!\!\!\perp_P$ means independence in probability.

In words, the potential outcomes under treatment are independent of treatment assignment. In this case, it can be ignored how units ended up in the treatment or control group. Equivalently, the group that received treatment could have been exchanged with the group receiving control, resulting in the same potential outcome.

The three assumptions together constitute the randomized controlled trial (as illustrated in Figure 3.2) and make calculation of the average treatment effect possible by means of reasoning with potential outcomes. Besides the use of potential outcomes, the potential outcome framework contains one additional element that enables one to bypass the fundamental problem of causal inference beyond randomized controlled trials, which is the assignment mechanism [118].

3.2.3 Beyond Randomized Controlled Trials

Unlike for randomized controlled trials, the ignorability assumption is easily violated when dealing with observational data, because the treatment and control groups are rarely truly exchangeable. A *confounder* can causally influence the treatment variable as well as the outcome variable as illustrated in Figure 3.3. Therefore, more lenient assumptions can be adopted to render the calculation of causal effects under the potential outcome framework still possible in the presence of confounders.

Assumption 4 (Conditional Ignorability)**.** Let $\boldsymbol{Z}$ denote confounding variables. Consider binary treatment assignment random variable T. Then conditional ignorability is satisfied if

$$Y(0), Y(1) \perp\!\!\!\perp_P T \mid \boldsymbol{Z}.$$

That means that the treatment and control group are generally not exchangeable, but they become exchangeable when conditioning on the confounding set. For that reason, *conditional ignorability* is also known as the *unconfoundedness* assumption. It

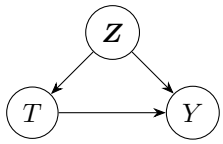


Figure 3.3: Because $\boldsymbol{Z}$ causally influences both T and Y, $\boldsymbol{Z}$ is said to *confound* the relation between T and Y.

is useful to adjust for confounding to reach conditional ignorability as long as each treatment level has non-zero probability within every subgroup defined by the confounders. The *positivity* assumption guarantees this condition holds.

Assumption 5 (Positivity)**.** Let $\boldsymbol{Z}$ denote confounding variables. Then positivity is satisfied if

$$P(T = t \mid \boldsymbol{Z}) \in (0, 1) \quad \forall\, T, \boldsymbol{Z}.$$

There is a tradeoff between conditional ignorability and positivity by virtue of adjusting for covariates [68], which is the process of conditioning on subgroups of the data that share similar covariate values. Intuitively, the more covariates are adjusted for, the smaller the subgroups become. This can lead to subgroups being entirely assigned to either treatment or control, which is a violation of the positivity assumption. Contrary, not sufficiently adjusting for high-dimensional covariates may lead to violations of conditional ignorability assumptions. Section 3.4.2 explains how this problem motivates the use of parametric approaches over non-parametric ones. Both conditional ignorability and positivity together are called *strong ignorability* [200, 119].

Vested with all of the above assumptions, one is able to calculate the average treatment effect. Assume binary treatment assignment variable T and confounding set $\boldsymbol{Z}$:

$$\begin{aligned}
\mathbb{E}[Y(1) - Y(0)] &\overset{(1)}{=} \mathbb{E}_{\boldsymbol{Z}}[\mathbb{E}[Y(1) - Y(0) \mid \boldsymbol{Z}]] \\
&\overset{(2)}{=} \mathbb{E}_{\boldsymbol{Z}}[\mathbb{E}[Y(1) \mid \boldsymbol{Z}] - \mathbb{E}[Y(0) \mid \boldsymbol{Z}]] \\
&\overset{(3)}{=} \mathbb{E}_{\boldsymbol{Z}}[\mathbb{E}[Y(1) \mid T = 1, \boldsymbol{Z}] - \mathbb{E}[Y(0) \mid T = 0, \boldsymbol{Z}]] \\
&\overset{(4)}{=} \mathbb{E}_{\boldsymbol{Z}}[\mathbb{E}[Y \mid T = 1, \boldsymbol{Z}] - \mathbb{E}[Y \mid T = 0, \boldsymbol{Z}]].
\end{aligned}$$

While the first two equalities follow from the laws of probability and expectation, the third equality is a result of conditional ignorability and positivity and the fourth equality a result of consistency. This result is also called the *adjustment formula* and the underlying assumptions are summarized in Figure 3.2. The formula requires one to have insight into the *assignment mechanism*: the conditional probabilities of treatment given covariates and potential outcomes. This is the second element that constitutes the potential outcome framework.

When conditional ignorability does not apply, causal inference becomes significantly harder. In some cases instrumental variables, those that causally influence the

treatment but not the outcome variable, can be utilized [93], the joint distribution of latent and observed confounders can be extracted from variational auto-encoders [147] and network data as a proxy for latent confounders can still be used to substantiate causal effects [89].

Consistency follows from the definitions of the structural causal models and hence the literature rejecting this assumption is not rich [178]. SUTVA can easily be violated by departures from the no-interference assumption. A framework that can account for such violations will be described; first, however, the structural causal model is introduced.

3.2.4 Structural Causal Models

Structural causal models (also known as structural equation models (SEMs)) are logically equivalent to the potential outcome framework [178] but offer a more formally structured notation:

Definition 3.5 (Structural Causal Models (SCM))**.** A structural causal model is a tuple $\mathcal{S} = (\boldsymbol{V}, \boldsymbol{W}, G, \boldsymbol{F})$ where $\boldsymbol{V} = \{V_1, \ldots, V_n\}$ is an ordered set of endogenous variables, $\boldsymbol{W}$ is a set of exogenous variables, G is a directed acyclic graph with vertex set $\boldsymbol{V}$, and $\boldsymbol{F} = \{f_1, \ldots, f_n\}$ is a set of structural functions satisfying:

1. For all $V_i \in \boldsymbol{V}$, there exist a corresponding subset of exogenous variables $\boldsymbol{W}_i \subseteq \boldsymbol{W}$ and a mapping $f_i : \Omega_{\mathbf{pa}(V_i) \cup \boldsymbol{W}_i} \to \Omega_{V_i}$ that maps the state space of endogenous parents of V_i together with $\boldsymbol{W}_i$ to the state space of V_i:

$$v_i = f_i(\boldsymbol{pa}_i, \boldsymbol{w}_i).$$

2. The exogenous variables $\boldsymbol{w}$ are drawn from a probability distribution $P(\boldsymbol{W})$ over the state space $\Omega_{\boldsymbol{W}}$.

SCMs can be either parametric or non-parametric. Non-parametric structural equation models are sometimes invoked because assumptions about functional forms between respective exogenous and endogenous variables are costly. It is important to note that the SCM does not assume the independence of exogenous variables.[2] However, when this additional property is satisfied, the models are known as non-parametric structural equation models with independent errors (NPSEM-ie) as will be

[2]In Definition 3.5 this can be observed from the fact that for $V_i, V_j \in \boldsymbol{V}$, the corresponding $\boldsymbol{W}_i$ and $\boldsymbol{W}_j$ can overlap.

elaborated on in Section 3.5. The SCMs are assumed to be acyclic, also called *recursive*. Recursiveness allows a topological sort to exist over the endogenous variables.

A natural extension of SEMs arises when the no-interference assumption is violated in the context of spatial spillover, where causal spillover occurs when changes in one unit influence outcomes in neighboring units through spatial interdependencies. This extension is known as *spatially explicit structural equation models* (SESEMs) [137]. This method merges the flexibility of structural equation models in describing complex relationships with the capability to explicitly model spatial confoundedness through variance/covariance matrices computed across various lag distances. An application of such a model will be discussed in Chapter 6.

Frequently, the true SEM is unattainable due to a limited ability to observe a system [201], and one has to settle for surrogate models that do not have equal expressive power, but can be sufficient to answer queries of the first two levels of the hierarchy. These surrogate models will be introduced in the following sections.

3.3 Associational Level

This section introduces concepts and associated assumptions necessary to address questions at the first level of Pearl's causal hierarchy, the associational level (see Figure 3.1). The section starts with some preliminaries on the relation between probability distributions and graphical models. It then explains the features of Bayesian networks and introduces Markov random fields. Because structure learning of Bayesian networks closely resembles causal discovery, Section 3.4.1 provides further information on structure learning. An overview of the items covered in this section is presented in Figure 3.4.

3.3.1 Bayesian Networks

In order to address queries at the first level, random variables need to be tied to the graphical components introduced. This is only possible when additional assumptions are invoked. Let $X_1, \ldots, X_n$ be random variables with joint probability distribution $P(x_1, \ldots, x_n)$. In *Bayesian networks* (BNs), the random variables are represented by the nodes of a directed acyclic graph and the probabilistic dependencies are represented by the edges via the *local Markov* assumption:

Assumption 6 (Local Markov)**.** Let $P(x_1, \ldots, x_n)$ be the joint probability distribution of random variables X_i corresponding to nodes $V_i \in \boldsymbol{V}$ in the directed acyclic

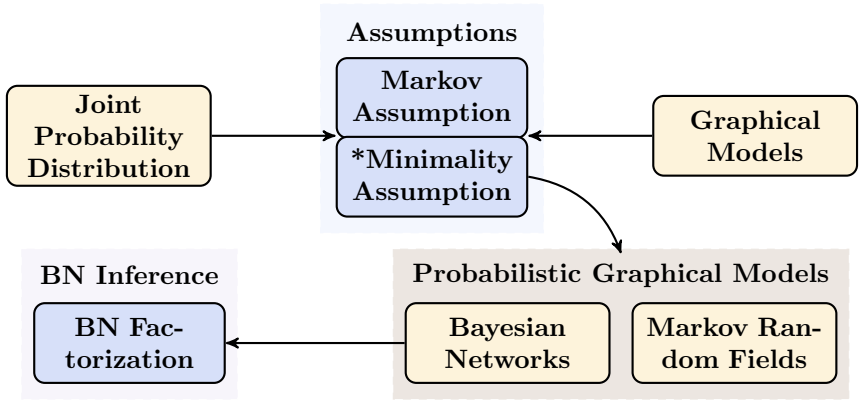


Figure 3.4: Assumptions (blue) and concepts (yellow) discussed at the associational level of the hierarchy. Probability distributions and graphical models can be tied together by means of the Markov assumptions. The minimality assumption can be adopted optionally for a parsimonious encoding of the joint distribution. The resulting object can either be a Bayesian network or a Markov random field. While inference is more extensively discussed in Chapter 5, the Bayesian network factorization assumption plays a central role.

graph $G = (\boldsymbol{V}, \boldsymbol{E})$. Then, the local Markov assumption holds if for every X_i the following holds in the graph:

$$X_i \perp\!\!\!\perp_P \mathbf{nonde}(X_i) \mid \mathbf{pa}(X_i).$$

Since the local Markov assumption ties the random variables together with the graphical structure, $\boldsymbol{V}$ is assumed to inherit all the probabilistic properties from $\boldsymbol{X}$. Henceforth, $P(v_1, \ldots, v_n)$ will be used instead of $P(x_1, \ldots, x_n)$ to denote the probability distribution of the random variables. The use of the underscore P to imply independence in probability is not superfluous as there also exists independence in the graph defined by *d-separation* and denoted by symbol $\perp\!\!\!\perp_G$.

Definition 3.6 (d-separation)**.** A path ρ between V_i and V_j (ignoring edge directions) is d-connected in the directed acyclic graph $G = (\boldsymbol{V}, \boldsymbol{E})$ by a set of nodes $\boldsymbol{C} \subseteq \boldsymbol{V} \backslash \{V_i, V_j\}$ if

1. ρ does not contain a chain $\cdots \to Z \to \cdots$ or fork $\cdots \leftarrow Z \to \cdots$, where Z is contained in $\boldsymbol{C}$.

2. all colliders (nodes where two arrows converge, e.g., $\cdots \to Z \leftarrow \cdots$) of the path ρ are in $\boldsymbol{C}$ or have a descendant in $\boldsymbol{C}$.

If there are no d-connecting paths between V_i and V_j given $\boldsymbol{C}$, then V_i and V_j are *d-separated* by $\boldsymbol{C}$ which is denoted by $V_i \perp\!\!\!\perp_G V_j \mid \boldsymbol{C}$.

The concept of graph independencies gives rise to a reformulation of the local Markov assumption to the global Markov assumption:

Assumption 7 (Global Markov)**.** Let $P(v_1, \ldots, v_n)$ be the joint probability distribution of random variables corresponding to the nodes $V_i \in \boldsymbol{V}$. Let $\perp\!\!\!\perp_G$ denote d-separation in the directed acyclic graph $G = (\boldsymbol{V}, \boldsymbol{E})$ and $\perp\!\!\!\perp_P$ independence in distribution. Then the global Markov assumption holds if for all $\boldsymbol{V}_I, \boldsymbol{V}_J, \boldsymbol{V}_K \subseteq \boldsymbol{V}$

$$\boldsymbol{V}_I \perp\!\!\!\perp_G \boldsymbol{V}_J \mid \boldsymbol{V}_K \implies \boldsymbol{V}_I \perp\!\!\!\perp_P \boldsymbol{V}_J \mid \boldsymbol{V}_K.$$

By relating the independencies of the graph to the independencies of the distribution, one can leverage the graphical structure for a parsimonious factorization of the joint probability distribution. This can also be directly assumed.

Assumption 8 (Bayesian Network Factorization)**.** Let $P(v_1, \ldots, v_n)$ be the joint probability distribution of random variables corresponding to the nodes $V_i \in \boldsymbol{V}$ in the directed acyclic graph $G = (\boldsymbol{V}, \boldsymbol{E})$. The Bayesian network factorization assumption holds if the distribution can be factorized according to the corresponding graphical structure:

$$P(v_1, \ldots v_n) = \prod_{i=1}^{n} P(v_i \mid \boldsymbol{pa}_i).$$

Example 1 (Graph Factorization)**.** Consider the Bayesian network displayed by Figure 3.3. According to the Bayesian network factorization assumption, the joint probability distribution $P(\boldsymbol{Z}, Y, T)$ can be factorized to $P(\boldsymbol{Z})P(T \mid \boldsymbol{Z})P(Y \mid \boldsymbol{Z}, T)$.

It has been shown that the local Markov assumption, the global Markov assumption and the Bayesian network factorization are equivalent when positivity is assumed [127]. When any of these equivalent conditions holds, the probability distribution P is said to be *Markov relative* (or Markov[3] in short) to $G = (\boldsymbol{V}, \boldsymbol{E})$.

While the Markov assumption imposes restrictions on the probability distribution via the graphical structure, an additional assumption is necessary to obtain a minimal representation of the probability distribution's conditional independence structure.

[3]The Markov assumption is in specific cases known as the causal Markov assumption. Technically, the assumption is only causal when the concomitant graphical component has causal meaning (which will be introduced in Section 3.4.2).

This assumption comes in various forms of increasing strength: SGS-minimality, P-minimality and faithfulness [263]. P-minimality will be discussed here [178], but before introducing this assumption, the concept of a *preferred* graph needs to be introduced:

Definition 3.7 (Preferred Graph)**.** Let P be the set of distributions that is Markov relative to $G = (\boldsymbol{V}, \boldsymbol{E})$ and $G' = (\boldsymbol{V}, \boldsymbol{E}')$. Then G' is (strictly) preferred to G if the conditional independence relations of G are a (proper) subset of the conditional independence relations of G'.

Assumption 9 (Minimality)**.** Let P be the set of distributions that is Markov relative to $G = (\boldsymbol{V}, \boldsymbol{E})$. Minimality is satisfied with respect to G if P is not Markov relative to a strictly preferred graph $G' = (\boldsymbol{V}, \boldsymbol{E}')$ to G.

Although minimality is a desirable assumption because it allows one to encode the joint distribution in the most parsimonious graphical structure possible, it is not required to answer queries at the first level of the hierarchy.

In concluding this section, not all independence relations are representable by a Bayesian network, as the following counterexample illustrates:

Example 2 (Limits of Bayesian Networks in Encoding Independencies)**.** Let X_1, X_2, X_3, X_4 be random variables. Then, there does not exist a Bayesian network satisfying conditional independence relations $X_1 \perp\!\!\!\perp_P X_2 \mid \{X_3, X_4\}$ and $X_3 \perp\!\!\!\perp_P X_4 \mid \{X_1, X_2\}$.

Therefore, there is another graphical structure that can represent conditional independencies: the *Markov random field*. Unlike Bayesian networks, which use directed edges to encode asymmetric relationships, Markov random fields use undirected edges to represent symmetric associations. This allows them to account for cyclic probability relations and work with potential functions, but prevents them from representing directionality. For more information about Markov random fields, the reader is referred to the work by Koller and Friedman [127]. Both Bayesian networks and Markov random fields are *probabilistic graphical models* as they unify joint probability distributions with graphical structures.

3.4 Interventional Level

This section discusses the causal assumptions and components necessary to address queries at the second level of the hierarchy. It begins with the various sets of assumptions required to conduct parametric as well as non-parametric causal discovery in Section 3.4.1, as specified in Figure 3.5. Section 3.4.2 then demonstrates how the

output of causal discovery, a causal diagram, forms the basis of both a non-parametric and a parametric approach, where the approaches differ based on a different appreciation of the fundamental problem of causal inference. The non-parametric approach adopts assumptions inherent to causal Bayesian networks that enable inference, while the parametric approach emerges by observing that the fundamental problem of causal inference requires estimation by definition. Figure 3.6 shows the specifications of the different concepts and assumptions necessary for causal inference for each of the two approaches. Finally, causal concepts that emerge when deviating from putative assumptions are discussed in Section 3.4.3.

3.4.1 Causal Discovery

This section will discuss causal discovery from the point of view of necessary assumption, expanding on previous assumptive approaches [69]. Technical details will be discussed when they are contingent on the introduced assumptions, but for a broader account of why causal discovery methods fail in the absence of assumptions, the reader is referred to a survey paper by Runge [207]. Although this section can serve as a blueprint for which method to use when certain assumptions are adopted, a more practical guide about the application of causal discovery methods can be found in the work by Malinsky and Danks [155]. While interventional data can significantly improve causal structure learning by resolving directional ambiguities that observational data cannot [96, 224], this section is restricted to recovering the structure with observational data alone. Because observational data alone is available at both the associational and interventional levels of the hierarchy (in the absence of actual interventions), structure learning at these two levels coincides [149]. Additionally, this section is limited to static causal discovery methods, which are causal discovery methods that do not account for the passage of time. There is a body of survey papers on causal discovery methods for longitudinal data and the additional assumptions necessary [17, 207].

An assumption most causal discovery methods revolve around is the *i.i.d.* assumption.

Assumption 10 (Independent and Identically Distributed (i.i.d.))**.** The observational data are independent and identically distributed.

Structure learning is first discussed under the assumptions of causal sufficiency, Markov, faithfulness, acyclicity, and independent and identically distributed data. Subsequently, causal discovery is considered in the presence of violations of the causal

sufficiency assumption, followed by a discussion of relaxations of the faithfulness assumption. Some of these approaches are summarized in Figure 3.5. However, there are assumption sets that allow conducting causal discovery beyond the assumption sets in Figure 3.5. Concepts that emerge when the Markov or the i.i.d. assumptions are violated are discussed in Section 3.4.3.

Because the goal of causal discovery is to recover as much of the graphical structure as possible from observational data, the core assumption within causal discovery should imply features of this underlying structure from the probability distributions (that are learned from the data). The strongest form of that assumption was already touched upon in Section 3.3.1 and is called *faithfulness*:

Assumption 11 (Faithfulness)**.** Let $P(v_1, \ldots, v_n)$ be the joint probability distribution of random variables $V_i \in \boldsymbol{V}$ corresponding to the nodes in the graph $G = (\boldsymbol{V}, \boldsymbol{E})$. Let $\perp\!\!\!\perp_G$ denote d-separation in a graph $G = (\boldsymbol{V}, \boldsymbol{E})$ and $\perp\!\!\!\perp_P$ be the independencies in distribution. Then the probability distribution P is faithful to G if for all $\boldsymbol{V}_I, \boldsymbol{V}_J, \boldsymbol{V}_K \subseteq \boldsymbol{V}$:

$$\boldsymbol{V}_I \perp\!\!\!\perp_P \boldsymbol{V}_J \mid \boldsymbol{V}_K \implies \boldsymbol{V}_I \perp\!\!\!\perp_G \boldsymbol{V}_J \mid \boldsymbol{V}_K.$$

A probability distribution can be faithful to a graph that is acyclic. If this is the case, then the *acyclicity* assumption holds in addition to faithfulness. Practitioners who adopt faithfulness are not necessarily expected to have access to the full probability distributions, but are equipped with appropriate independence tests to find (conditional) independencies in the data. In order to complete the first collection of assumptions necessary to conduct causal discovery, the *causal sufficiency* assumption is highlighted:

Assumption 12 (Causal Sufficiency)**.** A set of variables $\boldsymbol{V}$ is causally sufficient if there are no unobserved confounders, meaning $\boldsymbol{V}$ contains all common causes of any two or more variables in $\boldsymbol{V}$.

When causal sufficiency is assumed, the object to be construed is the directed acyclic graph that best fits the data generating process of the observational data. Because observational data can only identify conditional independence structures, multiple distinct graphical structures may be consistent with the same observational distribution. To represent this uncertainty, we introduce the following definition:

Definition 3.8 (Completed Partially Directed Acyclic Graph)**.** Directed acyclic graphs that entail the same conditional independencies are said to be in the same *Markov*

equivalence class for DAGs. The Markov equivalence class for DAGs is represented by a *completed partially directed acyclic graph* (CPDAG) for which an edge is directed if all directed acyclic graphs in the Markov equivalence class agree on the direction of the edge and undirected otherwise.

The causal sufficiency, Markov, faithfulness, acyclicity and i.i.d. assumptions make up the first assumption set that allows causal discovery.

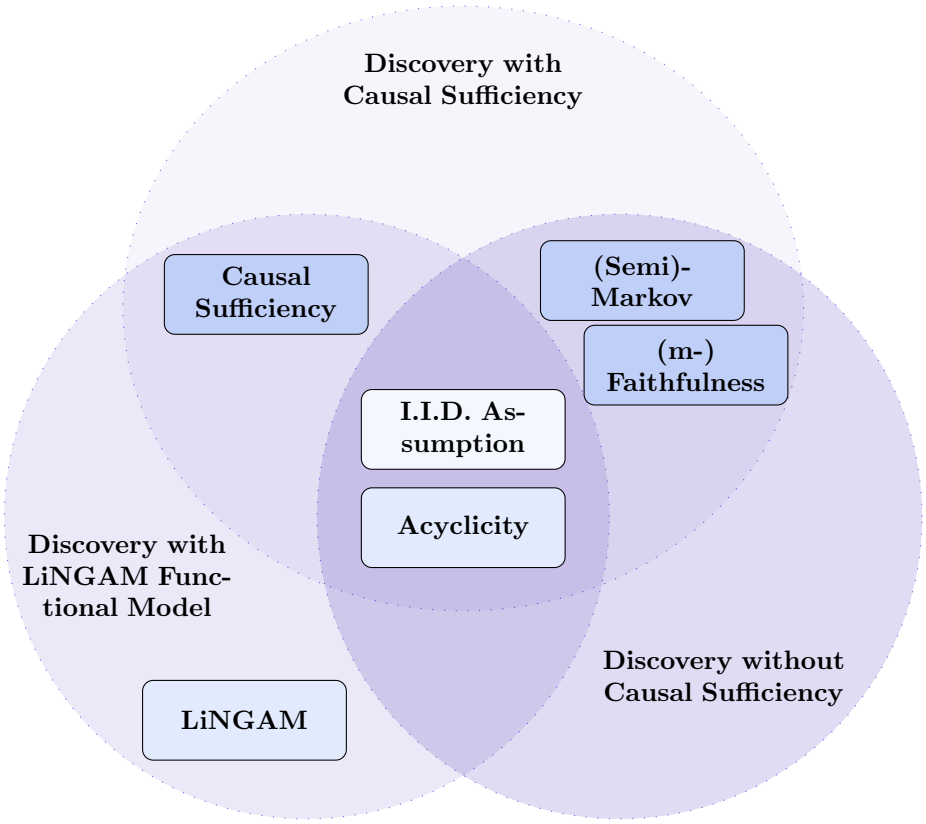


Figure 3.5: Causal discovery assumption sets: the different purple circles represent possible sets of assumptions described in Section 3.4.1 under which causal discovery can be conducted.[4] The boxes represent the assumptions necessary for causal discovery, which may have overlap with multiple assumption sets. The color of the boxes indicates the nature of the assumptions: while light blue represents sampling assumptions, ivory blue indicates assumptions on the data generating process, and darker blue is used for causal assumptions.

Causal discovery with causal sufficiency

Vested with this collection of assumptions as illustrated in the top circle of Figure 3.5, the structure of the underlying data generating process could be investigated with observational data alone. The first algorithm was the *Spirtes, Glymour and Scheines* algorithm [230], closely followed by the *Peter-Clarke* algorithm [229]. Both are *constraint-based* methods, meaning they aim to exploit the conditional independencies to inform the structure of the graph. This means that they require the use of reliable conditional independence testing methods. The algorithms output the CPDAG based on observational data.

Besides constraint-based methods, there are also *score-based* methods. Score-based methods employ the same assumptions, take in the same input, and generate the same output as constraint-based methods, but work fundamentally differently. The methods start with a specific CPDAG and fit it to the data. The fit is scored based on a scoring system and compared to the score of a slightly different CPDAG. The best fit is kept, and the algorithm continues in the same way. In order to restrain the enormous search space, they often have a forward and a backward phase. The forward phase keeps adding edges, which improves the score the most. When no edges can be added that can improve the score, the backward phase starts removing edges that improve the score the most. If there is no edge that can be removed to improve the score, the algorithm ends. Score-based methods require the use of the appropriate loss function based on the nature of the data. Common choices include the Bayesian information criterion, which balances model fit with complexity. An example of such a score-based method that has been proven to work well in simulation studies of small sample sizes is *greedy equivalent search* (GES) [48, 155].

Causal discovery without causal sufficiency

The causal sufficiency assumption can be relaxed. In this case, the possibility of missing common causes in the observational data is acknowledged, and the target of interest is expected to account for unobserved confounders. The smallest superclass of DAGs that accounts for the presence of unobserved confounders and is closed under marginalization is a *maximal ancestral graph* [193]. Similar to how multiple DAGs can encode the same independence constraints, multiple maximal ancestral graphs can also

[4]The list of assumption sets is not exhaustive as more possible assumption sets will be described that allow conducting causal discovery. Although constraint-based and score-based causal discovery algorithms require the use of appropriate conditional independence tests and scoring methods respectively, these are not mentioned as assumptions because they are algorithm-specific.

represent the same conditional independencies. This gives rise to the *partial ancestral graph* that represents the Markov equivalence class of maximal ancestral graphs with the same independence constraints.

It is important to note that the existence of unobserved confounding also leads to a slightly modified version of d-separation that represents conditional independencies with respect to the maximal ancestral graph, called *m-separation.* This leads to natural extensions of the Markov assumption and the faithfulness assumption that go by the *semi-Markov* assumption and *m-faithfulness.*

Algorithms that can extract the partial ancestral graph from observational data such as *fast causal inference* [231], *greedy fast causal inference* [171] and *really fast causal inference* [52] rely on the i.i.d. assumption, the semi-Markov assumption and the m-faithfulness assumption to an acyclic system as illustrated in the bottom right circle of Figure 3.5.

There are two main drawbacks with the algorithms introduced so far. First, either traditional faithfulness or its extension to unobserved confounder models (m-faithfulness) is assumed. Faithfulness is a strong assumption, and it can be easy to find examples where faithfulness is violated [7]. Second, the output of all introduced algorithms entails a representation of a Markov equivalence class. In order to exploit the obtained graphical structure for inference purposes, additional assumptions on the data generating process should be adopted to direct the edges in the graphical structure, which the algorithm could not provide. Both drawbacks can be skirted by assuming restrictions on the data generating process beforehand. This will be discussed in the next section.

Parametric causal discovery and relaxations of faithfulness

In Pearl's causal hierarchy, the true object of investigation is the structural causal model. Because the true SCM is almost always unattainable, one is forced to settle for a surrogate model for which at least questions of lower levels of the hierarchy can be addressed. However, by taking parametric assumptions on the distribution of the underlying SCM, other assumptions can be bypassed.

These methods are based on *functional causal models*, which are equivalent to earlier introduced SCMs [86], where one writes the dependent variable as a function of its parents and a noise term. A special case of a functional causal model is *linear non-Gaussian acyclic model* (LiNGAM) and is defined as follows:

Assumption 13 (LiNGAM)**.** A SCM $\mathcal{S}$ with an ordered set of endogenous variables

$\boldsymbol{V} = \{V_1, \ldots, V_n\}$, exogenous variables $\boldsymbol{W} = \{W_1, \ldots, W_n\}$ and a set of functions $\boldsymbol{F} = \{f_1, \ldots, f_n\}$ is assumed to be a linear non-Gaussian acyclic model if:

1. Every endogenous variable v_i is a linear combination of its parents in the topological sort and exogenous variable term w_i:

$$v_i = f_i(\boldsymbol{pa}_i, w_i) = \sum_{j:V_j \in \mathbf{pa}(V_i)} b_{ij} v_j + w_i.$$

2. The error terms w_i are drawn from exogenous variables $W_i \in \boldsymbol{W}$, which are continuous, mutually independent, and follow a non-Gaussian distribution.

When LiNGAM is assumed, methods exist to fully recover the DAG based on independent component analysis (ICA-LiNGAM) [218]. Faithfulness can be dropped, but causal sufficiency, acyclity and the i.i.d. assumptions should be adopted. The assumption set has been summarized in Figure 3.5. Complementary LiNGAM discovery methods were further developed to account for the violation of causal sufficiency [110]. In addition, there are also variants that allow for a violation of the acyclicity assumption [134].

There are also alternative assumptions (to LiNGAM) on the data generating process that can be used to sideline the faithfulness assumptions and retrieve the full DAG. Some of those assume an additive noise data generating process [109, 185]. More general methods assume a post-linear form [266], where it has been proven that in all but 5 model specification cases the causal direction is identifiable. Even though faithfulness does not have to be assumed in some cases, less restrictive assumptions do have to be adopted [185].

If one is not willing to commit to additional assumptions about the data generating process, but still considers faithfulness too strong of an assumption, one can adopt one of the many weaker versions of faithfulness [265], such as adjacency faithfulness [231, 189], 2-adjacency faithfulness [157] and frugality [74] for which causal discovery algorithms exist or could be developed.

Finally, research has shown that causal discovery algorithms can be unstable [126] or that only limited parts of the graphical structure can be discovered from pure observations [184]. For this purpose, domain knowledge can be used to refine the performance of causal discovery algorithms and has been incorporated via tiered background knowledge [10], user interactions [152], or the penalization of the search process [94]. It is recommended that practitioners assess the possibility of incorporating domain knowledge or experts to enhance the quality of the obtained graphical structure.

3.4.2 Identification and Inference

This section discusses how the concepts from causal discovery can be used for parametric as well as non-parametric inference. While the debate regarding the extent to which the result of causal discovery can be termed 'causal' is acknowledged [62], this section assumes that the ADMGs and DAGs convey causal meaning, making them *causal diagrams*. It first addresses how non-parametric causal inference has contributed to causal inference and emphasizes its assumptions. Subsequently, the assumptions adopted by the parametric approach to causal inference are described, along with points of convergence between the two approaches. Both approaches are summarized in Figure 3.6.

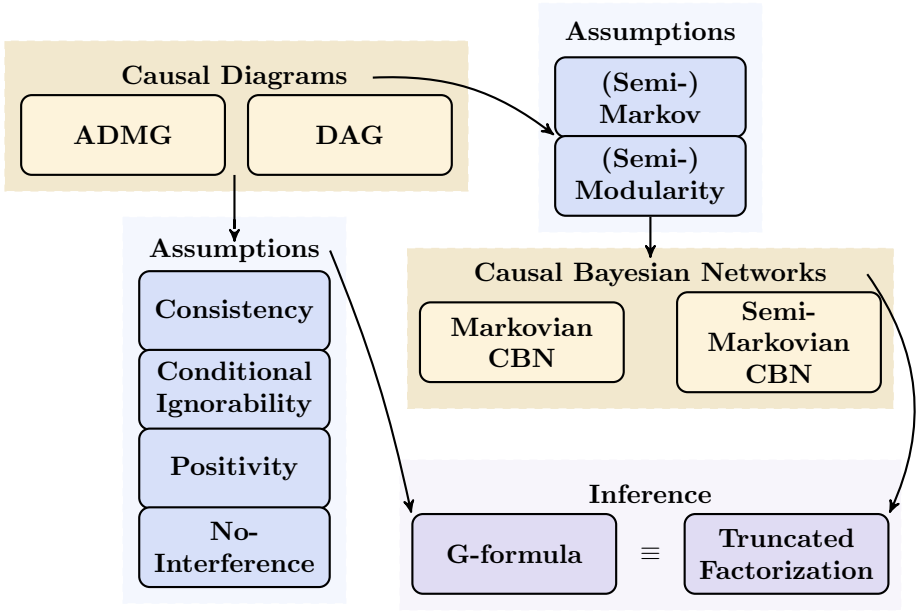


Figure 3.6: Causal diagrams are the basis for causal inference. They can be endowed with assumptions from Section 3.2 to allow inferring causal statements under the g-formula. Alternatively, the diagrams can be subjected to non-parametric assumptions to obtain causal Bayesian networks, which can be leveraged for inference with the truncated factorization formula.

Non-parametric causal inference

In order to infer causal statements, the causal meaning should be specified on top of the Bayesian networks that were introduced earlier. This leads to the definition of

causal Bayesian networks. The 'missing link' definition, as described by Bareinboim et al. [26], is adopted among multiple equivalent definitions of causal Bayesian networks, as it intuitively implicates the (SGS-)minimality assumption. The assumptions inherent to the definitions are examined. Central to this notation are (atomic) interventions; therefore, the *do*-operator and the associated interventional distribution are introduced.

Definition 3.9 (Interventional Distribution)**.** Let Y be a random variable and $\boldsymbol{S} \subset \boldsymbol{V}$ be a set of random variables. The interventional distribution $P(y \mid do(\boldsymbol{S} = \boldsymbol{s}))$ encodes the probability that $Y = y$ given that $\boldsymbol{S}$ is forced to take value $\boldsymbol{s}$ (denoted by the *do-operator* $do(\boldsymbol{S} = \boldsymbol{s})$, or $do(\boldsymbol{s})$ in short) with probability 1.

Bayesian networks that do not contain latent variables are first considered; these are referred to as *Markovian*.

Markovian causal Bayesian networks: The behavior of the do-operator within a Bayesian network can be assumed by the modularity assumption.

Assumption 14 (Modularity)**.** Let P be a probability distribution Markov relative to Bayesian network $G = (\boldsymbol{V}, \boldsymbol{E})$ and let $\boldsymbol{S} \subseteq \boldsymbol{V}$. Then an intervention $do(\boldsymbol{S} = \boldsymbol{s})$ is said to be modular if:

1. For every $V_i \in \boldsymbol{V} \backslash \boldsymbol{S}$, where $\boldsymbol{S}$ and $\mathbf{pa}(V_i)$ are disjoint in G, the interventional distribution by intervening on the parents of V_i is invariant to other interventions in the graph:

$$P(v_i \mid do(\boldsymbol{S} = \boldsymbol{s}), do(\mathbf{pa}(V_i) = \boldsymbol{pa}_i)) = P(v_i \mid do(\mathbf{pa}(V_i) = \boldsymbol{pa}_i)).$$

2. For every $V_i \in \boldsymbol{V}$, the interventional distribution by intervening on the parents of V_i yields the same distribution as observing the parents of V_i:

$$P(v_i \mid do(\boldsymbol{S} = \boldsymbol{s}), do(\mathbf{pa}(V_i) = \boldsymbol{pa}_i)) = P(v_i \mid do(\boldsymbol{S} = \boldsymbol{s}), \mathbf{pa}(V_i) = \boldsymbol{pa}_i).$$

Modularity specifies how the interventional distributions operate within the context of a Bayesian network. A *causal Bayesian network* can now be defined:

Definition 3.10 ((Markovian) Causal Bayesian Network (CBN))**.** Let P be a probability distribution Markov relative to Bayesian network $G = (\boldsymbol{V}, \boldsymbol{E})$. Then $G = (\boldsymbol{V}, \boldsymbol{E})$ is said to be a causal Bayesian network if for all $\boldsymbol{S} \subseteq \boldsymbol{V}$ and $V_i \in \boldsymbol{V} \backslash \boldsymbol{S}$:

1. $P(v_i \mid do(\boldsymbol{S} = \boldsymbol{s}))$ is Markov relative to G.

2. The intervention $do(\boldsymbol{S} = \boldsymbol{s})$ is modular.

The assumptions of the interventional distributions implicit in the definition of causal Bayesian networks immediately imply (SGS-)minimality in case the conditional probability distributions are strictly positive. In case they are deterministic, there still is good reason to assume (SGS-)minimality [264].

As the Markov assumption implies a factorization of a Bayesian network, in a similar way the modularity assumption implicit in the causal Bayesian networks enforces the *truncated factorization* for interventional distributions [26]:

Assumption 15 (Truncated Factorization)**.** Let P be a probability distribution Markov relative to Bayesian network $G = (\boldsymbol{V}, \boldsymbol{E})$. Let $\boldsymbol{S} \subseteq \boldsymbol{V}$ be the set of random variables where is intervened upon. The truncated factorization is assumed to hold if:

$$P(\boldsymbol{v} \mid do(\boldsymbol{S} = \boldsymbol{s})) = \prod_{i|V_i \notin \boldsymbol{S}} P(v_i \mid \boldsymbol{pa}_i) \quad \text{if } \boldsymbol{v} \text{ consistent with intervention } \boldsymbol{s}$$

and 0 otherwise.

The truncated factorization property implicit in Markovian causal Bayesian networks reduces marginal inference in Markovian causal Bayesian networks to marginal inference in the *mutilated Bayesian networks*. These are the networks that are obtained when removing all the arrows to these nodes where is intervened upon. Although the truncated factorization property is sometimes known as the *g-formula* [182], it will be emphasized in Section 3.4.2 that the g-formula is derived from a different appreciation of the fundamental problem of causal inference as shown in Figure 3.6.

More efficiently, the interventional distribution can be computed by means of the adjustment set and associated adjustment formula for causal Bayesian networks [244]:

Definition 3.11 (Adjustment Set)**.** Let P be a probability distribution Markov relative to Bayesian network $G = (\boldsymbol{V}, \boldsymbol{E})$. An adjustment set is a set $\boldsymbol{V}_J \subset \boldsymbol{V}$ for which:

$$P(\boldsymbol{v}_S \mid do(\boldsymbol{V}_K = \boldsymbol{v}_K)) = \begin{cases} P(\boldsymbol{v}_S \mid \boldsymbol{v}_K) & \text{if } \boldsymbol{V}_J = \emptyset, \\ \sum_{\boldsymbol{v}_J} P(\boldsymbol{v}_S \mid \boldsymbol{v}_K, \boldsymbol{v}_J) P(\boldsymbol{v}_J) & \text{otherwise.} \end{cases} \tag{3.2}$$

Semi-Markovian causal Bayesian networks: The concepts and assumptions introduced in this section do naturally extend to the case when the models allow for unobserved confounding variables, as is the case in *semi-Markovian* models. Naturally, the Markov assumption cannot be adopted but is replaced by a semi-Markov assumption. Although the full specifications of the semi-Markovian causal Bayesian network have been detailed by Bareinboim et al. [27], it is important to emphasize that inherent to that definition is an adjusted version of the Markov assumption and modularity assumption, tailor-made to account for the complexities when latent variables are involved.

As described in Section 3.4.1, the object that emerges when unobserved confounding random variables are at play is an ADMG. Naturally, the Markov assumption as defined above does not hold when unobserved confounders are involved, because the latent confounders cannot be conditioned on. By generalizing d-separation to m-separation, the Markov assumption can be extended to ADMGs [194], resulting in the semi-Markov assumption. Similarly, as in the original Markov assumption, the semi-Markov assumption can also be expressed in terms of m-separation or in terms of the truncated factorization of the distribution. It has been shown that both definitions are equivalent [194], but for specifications of the semi-Markov assumption or the associated semi-modularity assumption, the reader is referred to the article by Bareinboim et al. [27]. These assumptions together give rise to the *semi-Markovian causal Bayesian network*

Definition 3.12 (Semi-Markovian Causal Bayesian Network)**.** Let P be a probability distribution Markov defined on the ADMG $G = (\boldsymbol{V}, \boldsymbol{E})$. Then $G = (\boldsymbol{V}, \boldsymbol{E})$ is said to be a semi-Markovian causal Bayesian network if for all $\boldsymbol{S} \subseteq \boldsymbol{V}$ and $V_i \in \boldsymbol{V} \backslash \boldsymbol{S}$:

1. $P(v_i \mid do(\boldsymbol{S} = \boldsymbol{s}))$ is semi-Markov relative to $G_{\overline{\boldsymbol{S}}}$.

2. The intervention $do(\boldsymbol{S} = \boldsymbol{s})$ is semi-modular.

Obviously, the factorization implied by the semi-Markov assumption also leads to a form of truncated factorization of interventional distributions. For a full overview of this factorization and subsequent ways to marginalize out variables, the reader is referred to (the appendix of) Bareinboim et al. [27]. It has been proven that the do-calculus provides a complete toolkit necessary to rewrite interventional distributions to observational distributions, and the rules of do-calculus are implied by the assumptions implicit in the definition of the semi-Markovian Bayesian network [220]. Completeness of the do-calculus means that the do-calculus will provide an observational distribution

for each interventional distribution if it exists. When the interventional distributions cannot be written in observational terms, the distribution is called *unidentifiable*. Identification is a necessary condition for both non-parametric and parametric causal inference approaches

Parametric causal inference

Apart from some causal discovery methods, most of the concepts discussed so far are non-parametric concepts. Since potential outcomes by nature imply missing values, the fundamental problem of causal inference is essentially an *estimation problem*. That is why substantial contributions to causal inference also involve estimation. The motivation for parametric causal inference is briefly discussed, followed by an examination of the parametric counterpart of the truncated factorization (parametric g-formula), based on assumptions introduced in Section 3.2. At the third level of the hierarchy, these concepts will be extended (see Section 3.5).

The following example motivates the use of parametric methods as a result of estimation problems: according to the adjustment formula, the interventional probability $P(y \mid do(T = t))$ corresponding to the DAG of Figure 3.3 can be converted to observation probabilities:

$$P(y \mid do(T = t)) = \sum_{\boldsymbol{z}} P(y \mid T = t, \boldsymbol{z}) P(\boldsymbol{z}).$$

This is also known as the back-door adjustment [178]. Although using parametric methods would require additional assumptions on the functional form, there are two main benefits to using parametric approaches. First, when considering continuous treatment variables, the query of interest $P(y \mid do(T = t))$ might not be available from data for the intervention $do(T = t)$ of interest. Second, taking into account high-dimensional covariates $\boldsymbol{Z}$, summing over all the strata $\boldsymbol{z}$ could be intractable. Both estimation problems can be circumvented by assuming the functional form [100].

When returning to the fundamental problem of causal inference and the adjustment formula as a result of various assumptions in Section 3.2, calculating the conditional expectation $\mathbb{E}[Y \mid do(T = t)]$ of Figure 3.3 can be reduced to evaluating $\mathbb{E}_{\boldsymbol{Z}}\mathbb{E}[Y \mid T, \boldsymbol{Z}]]$. This would require the evaluation of $\mathbb{E}[Y \mid T, \boldsymbol{Z}]$ adjusted for the probability $P(\boldsymbol{z})$. However, a non-parametric evaluation of $\mathbb{E}[Y \mid T, \boldsymbol{Z}]$ is impossible when $\boldsymbol{Z}$ is high-dimensional. Therefore, one can fit a regression model to the data to receive the estimates for $\mathbb{E}[Y \mid T, \boldsymbol{Z}]$ for each combination of $(t, \boldsymbol{z})$ and only estimate the $P(\boldsymbol{z})$ for the $\boldsymbol{z}$ that are present in the data. This is called *standardization based on parametric*

models, or in a more general form, *the parametric g-formula*.

Alternatively, $\mathbb{E}[Y \mid do(T = t)]$ can be further reduced to

$$\mathbb{E}[Y \mid do(t)] = \sum_{y} \sum_{\boldsymbol{z}} \frac{yP(y, t, \boldsymbol{z})}{P(t \mid \boldsymbol{z})},$$

meaning $\boldsymbol{z}$ can be marginalized out from the joint probability if the conditional probability $P(t \mid \boldsymbol{z})$ for ending up in the treatment group $T = t$ is taken into account. When $\boldsymbol{Z}$ is high-dimensional, this cannot be completed with non-parametric methods, but parametric model specifications need to be assumed. Logistic regression would be a straightforward choice in case of binary treatment. This is an example of *inverse probability weighting*.

Together with g-estimation methods, inverse probability weighting and the parametric g-formula belong to the family of *g-methods*, a class of methods that allows the computation of the average causal effects under time-varying treatments [166]. All these methods rely on the availability of a causal diagram and on assumptions that have been described in Section 3.2. These assumptions include consistency, positivity, (conditional) ignorability, and no-interference as illustrated by Figure 3.6. The connection between the g-formula and the truncated factorization formula looms large because the latter stems from the non-parametric causality research, while the former originates in its parametric counterpart, both being derived from different assumptions.

In a similar way, expressions with the do-operator, such as $\mathbb{E}[Y \mid do(T = t)]$, can be formulated as expressions containing potential outcomes, $\mathbb{E}[Y(t)]$. Nonetheless, identifiable potential outcomes queries cannot always be reduced to observational queries via the do-calculus, as nested counterfactuals require more refined tooling for reduction. Section 3.5 explains that some properties of the do-calculus can be extended to account for the reduction of nested counterfactuals to observational queries as well [156].

3.4.3 Discovery, Identification and Inference with More Relaxations

There are also many more departures from traditional assumptions in causal discovery and inference that have been omitted so far and will be discussed here. Deviations that are henceforth considered are departures from the no-interference assumption, departures that allow for context-specific independence and departures that consider a different kind of intervention.

All of the discussed causal discovery methods in Section 3.4.1 are based on the i.i.d. assumption as illustrated by Figure 3.5. There is also an entire body of work in terms of causal discovery and inference when this assumption is violated [13, 151, 150, 139, 113, 239, 172, 180, 33, 217, 15]. As has been tenaciously demonstrated, the graphical structures that emerge as a result of causal discovery under interference, depend on the different kinds of causal interference present [172]. Causal research under interference has been bifurcating.

On the one hand, graphs with violations of the i.i.d. assumption allow directed edges for causal relationships as well as undirected edges for stable symmetric relationships. These can consequently be accounted for by either *Laurritzen-Wermuth-Frydenberg chain graphs* [138, 139, 33] or *Andersson-Madigan-Perlman chain graphs* [8] depending on the Markov property interpreted. Generalization of the former by relaxing causal sufficiency leads to segregated graphs [219, 217]. Complete identification and inference methods for segregated graphs with stable symmetric relationships are established [217]. Alternatively, an absorption of the Andersson-Madigan-Perlman chain graphs in combination with ADMGs [193] leads to a new family of graphical structures for which causal discovery methods exist for observational and interventional data [180].

On the other hand, extending the rules of d-separation to *relational d-separation*, a criterion for conditional independence in case of relational data, has given rise to an alternative representation, that enables the existence of independencies of relational data, called *abstract ground graphs* [150]. With an extension of the Peter-Clark algorithm, the *relational causal discovery* algorithm [151] makes it possible to extract the true relational causal structure in case of violations of the no-interference assumption. For every perspective, the relational causal model corresponds to an abstract ground graph. Inference is also possible under abstract ground graphs [13].

Because the Markov assumption has occasionally been defended [97] and criticized [43], there have also been attempts to relax the Markov assumption. Claiming that any variable is independent of its non-descendants given its parents excludes the possibility of conditional independence relations that only hold for a subset of realizations of conditioning variables [67]. Relaxing the Markov property to a kind of Markov property that allows for *context-specific independence* relations calls for different causal concepts that can account for this such as Bayesian multinets [80], conditional probability tables with regularity structure [37], staged trees and chain event graphs [226] and labeled directed acyclic graphs [181]. Various algorithms for causal discovery exist for staged trees [42, 142] as well as for labeled directed acyclic graphs [115] (with a

slightly adapted version of faithfulness). There are also inference methods available when context-specific independence is involved [240].

Besides the atomic or hard interventions discussed in Section 3.4.2, there are also stochastic or soft interventions. These interventions do not force the intervened variable to take on a fixed value, but merely replace the underlying causal mechanism by a known function [53, 70]. The do-calculus falls short in converting causal queries with soft interventions or conditional interventions. For that, a more general calculus is required, called *σ-calculus* [53] that can account for stochastic interventions and comes with a concomitant inference algorithm.

3.4.4 Practical Guide to Causal Inference

In this section, three practical considerations for conducting causal inference are discussed. Figure 3.7 summarizes key considerations for converting observational data into causal information, outlining the flow from data to causal graphical structures, estimands, and estimates, while highlighting the role of assumptions and expert knowledge.

First, identifying the necessary components to address inference queries of interest

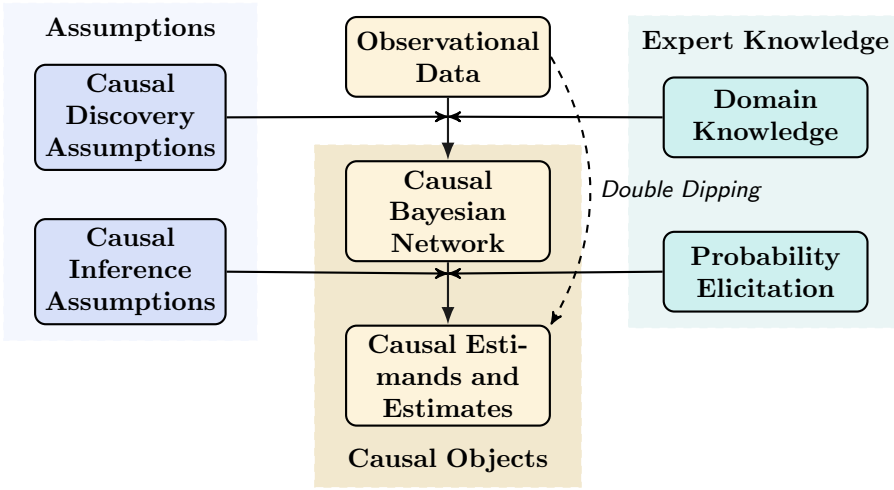


Figure 3.7: Flow of causal reasoning from data to graphical components (causal discovery) and subsequently to causal estimands and estimates (causal inference). While the light blue part indicates at what stage assumptions have to be taken into account, the light green part indicates possible supplementary information from domain experts. Double dipping occurs when data in the causal discovery stage is being reused at the causal inference stage.

requires a sound graphical structure. This graphical structure can be obtained via domain experts or causal discovery methods. When causal inference methods are applied to data that has already undergone causal discovery algorithms, this can result in *double dipping*, compromising the validity of the confidence intervals provided by the statistical methods [47]. Practitioners should be mindful of this issue and, if needed, apply available methods that can correct for this bias [87].

Second, it can be the case that there is no data available to estimate the conditional probability distributions (or one does not want to engage in double dipping). In this case, practitioners can still engage in causal inference when the conditional probability distributions are elicited from domain experts. There already exist efficient methods to infer discrete conditional probability distributions from experts [29, 5, 95], but eliciting continuous conditional probability distributions remains underdeveloped.

Finally, as discussed, causal inference relies on statistical methods, where larger sample sizes improve reliability. However, when $\boldsymbol{Z}$ is high-dimensional, non-parametric estimation of $\mathbb{E}[Y \mid T, \boldsymbol{Z}]$ becomes infeasible [100]. To address this, parametric methods such as the *parametric g-formula* and commonly parametric implementations of inverse probability weighting can be used, though they require specifying functional forms. When these assumptions are restrictive, semi-parametric alternatives offer a balance between flexibility and efficiency [34, 148].

3.5 Counterfactual Level

The components introduced in the previous two sections are not sufficient to address queries at the third level of the hierarchy. While the second level represents interventions on conditioning variables, the third level corresponds to interventions on conditioned variables. As mentioned in Section 3.2.4, the object necessary to reason at all levels of the hierarchy, including the counterfactual level, is the SCM. Next is an example of how an SCM can be utilized to reason at the counterfactual level of the hierarchy when the causal Bayesian network falls short:

Example 3 (Counterfactual Queries Require SCMs)**.** Assume the linear Gaussian (Markovian) causal Bayesian network corresponding to the graph $X \to Y$ with

$$\begin{aligned} X &\sim \mathcal{N}(1, 4) \\ Y &\sim \mathcal{N}(-0.5X + 3, 1). \end{aligned}$$

The intervention distribution $P(Y \mid do(X = 1))$ can be computed via the truncated

factorization formula and results in $\mathcal{N}(2.5, 1)$. However, the counterfactual distribution $P(Y(X = 0) \mid X = 1, Y = 4)$, meaning the probability of Y had X been set to 0 given that $X = 1$ and $Y = 4$, cannot be computed with a causal Bayesian network alone. In order to compute this counterfactual query, access to the SCM is required.

Therefore, assume the following structural equations in the SCM:

$$\begin{aligned} f_1(w_1) &= w_1 && \text{where } w_1 \sim \mathcal{N}(1, 4) \\ f_2(X, w_2) &= -0.5X + w_2 && \text{where } w_2 \sim \mathcal{N}(3, 1). \end{aligned}$$

The evidence of the counterfactual query, $X = 1$ and $Y = 4$, can be used to update the distribution of the exogenous variables in the SCM to $w_1 \sim \delta(1)$ and $w_2 \sim \delta(4)$, with $\delta()$ being the Dirac delta measure. Ingesting the intervention $X = 0$ into the updated structural equations leads to a complete evaluation of the counterfactual query: $P(Y(X = 0) \mid X = 1, Y = 4) = f_2(X = 0, w_2) = \delta(4)$.

One of the reasons much research has been dedicated to the first two levels of the hierarchy is that access to the fully specified SCM is considered to be implausible. While the above linear Gaussian (Markovian) Bayesian network gives rise to a natural separation between the endogenous and exogenous variables, the interaction between the observed and latent variables is often unknown, rendering access to the fully specified SCM 'hopeless' [27]. Despite the inaccessibility of the fully specified SCM, scholars have painstakingly reasoned with counterfactual models, because it plays an essential role in mediation analysis [197, 196]. Some counterfactual models have antagonized scholars that have argued that the introduced assumptions are not scientific because they lack the possibility of empirical validation [61].

This section generalizes the potential outcome framework introduced in Section 3.2.1, which is equivalent to the structural causal model framework presented in Section 3.2.4, thereby shedding new light on the assumptions involved at the third level of the hierarchy (see Figure 3.8). The different counterfactual models emerging from assumptions are emphasized, and the inference tools available for each model are highlighted. Throughout this section, the existence of a topological sort on the random variables is assumed.

[5]The SWIG does not immediately apply the edge g-formula, but the graphical structure of the SWIG can be generalized to allow edge interventions [222].

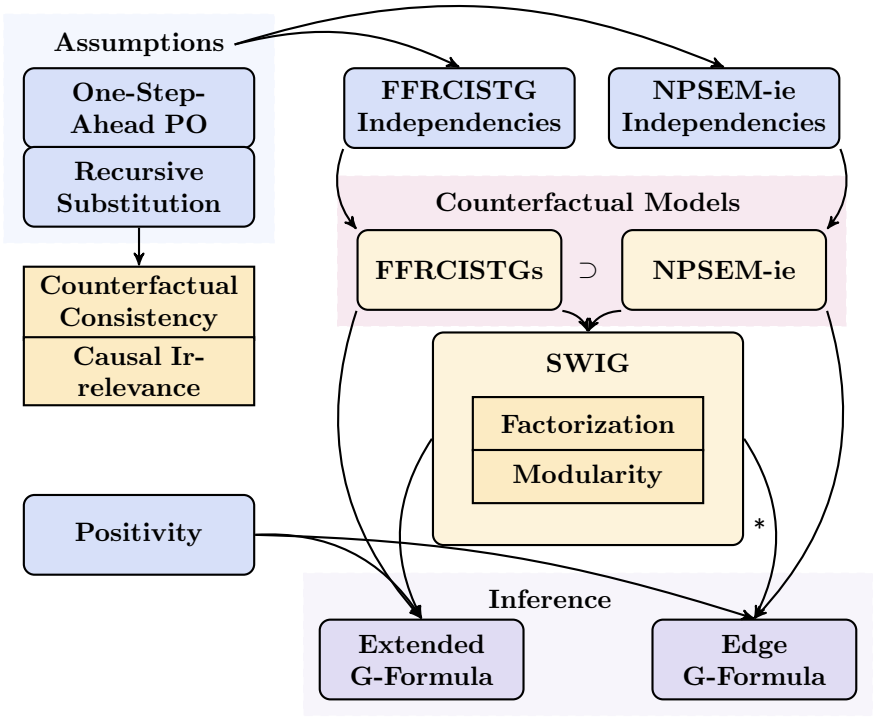


Figure 3.8: The definition of the one-step-ahead potential outcomes and recursive substitution imply the desirable counterfactual consistency and causal irrelevance property. Additional independence assumptions need to be adopted to yield a counterfactual model, which can either be a FFRCISTG or a NPSEM-ie. The SWIG unifies these models with graphical approaches and features a factorization and modularity property. Together with the positivity assumption, inference can be conducted via the extended g-formula or the edge g-formula.[5]

3.5.1 One-Step-Ahead Potential Outcomes

The very definition of counterfactuals entails the existence of a hypothetical world that may not be empirically verifiable. Therefore, the existence of *one-step-ahead potential outcomes* is assumed.

Assumption 16 (One-Step-Ahead Potential Outcomes)**.** Let $X_1, \dots, X_n$ be random variables corresponding to nodes $V_1, \dots, V_n$. Then for all $V_i \in \boldsymbol{V}$ and possible assignments of parents $\boldsymbol{pa}_i \in \Omega_{\mathbf{pa}(V_i)}$, the existence of one-step-ahead potential outcomes $V_i(\mathbf{pa}(V_i) = \boldsymbol{pa}_i)$ is assumed.

Note that $V_i(\mathbf{pa}(V_i) = \boldsymbol{pa}_i)$ corresponds to the notation introduced in the potential outcome framework of Section 3.2.1. Intuitively, the one-step-ahead potential

outcome corresponds to the response V_i had the parents of V_i been set to $\boldsymbol{pa}_i$. This is emphasized as an assumption because the assumed potential outcomes could possibly be counterfactual and therefore presuming the existence of a hypothetical world. Since not all potential outcomes naturally depend on possible assignments of parent nodes in the topological sort, it is necessary to extend the definition of potential outcomes via recursive substitution.

Assumption 17 (Recursive Substitution)**.** Let $\boldsymbol{X} = \{X_1, \ldots, X_n\}$ be random variables corresponding to nodes $\boldsymbol{V} = \{V_1, \ldots, V_n\}$. Assume the existence of one-step-ahead potential outcomes $V_i(\mathbf{pa}(V_i) = \boldsymbol{pa}_i)$ for all $V_i \in \boldsymbol{V}$ and possible assignments of parents $\boldsymbol{pa}_i \in \Omega_{\mathbf{pa}(V_i)}$. Then for all $\boldsymbol{S} \subset \boldsymbol{V}$ and $\boldsymbol{s} \in \Omega_{\boldsymbol{S}}$ it is assumed that $V_i(\boldsymbol{s})$ can be expressed recursively:

$$V_i(\boldsymbol{s}) = V_i(\boldsymbol{s} \cap \boldsymbol{pa}_i, \{V_j(\boldsymbol{s}) \mid V_j \in \mathbf{pa}(V_i), V_j \notin \boldsymbol{S}\}).$$

$V_i(\boldsymbol{s})$ is thus the potential outcome where the parents of V_i that are in $\boldsymbol{S}$ had been set to $\boldsymbol{s}$ and variables for which $V_j \in \mathbf{pa}(V_i) \setminus \boldsymbol{S}$ are set to the values these potential outcomes would have had had $\boldsymbol{S}$ been set to $\boldsymbol{s}$, denoted by $V_j(\boldsymbol{s})$.

Example 4 (Recursive Substitution in Graphical Strucures)**.** Assume the topological sort over the random variables $\boldsymbol{Z}, T, Y$ as implied by Figure 3.3. Then, it is assumed that the one-step-ahead potential outcome $Y(\boldsymbol{z})$ is defined recursively:

$$\begin{aligned} Y(\boldsymbol{z}) &= Y(\boldsymbol{z} \cap \boldsymbol{pa}_Y, \{V_j(\boldsymbol{z}) \mid V_j \in \mathbf{pa}(Y), V_j \notin \boldsymbol{Z}\}) \\ &= Y(\boldsymbol{z}, T(\boldsymbol{z})). \end{aligned}$$

Expressing potential outcomes recursively brings along desirable properties as illustrated by Figure 3.8. First of all, it directly implies the consistency assumption introduced in Section 3.2 [156]. Second, it proves the so-called *causal irrelevance*: every potential outcome derived from recursive substitution $V_i(s)$ can be expressed as a unique minimally causal relevant subset of $W \subseteq S$: $V_i(s) = V_i(w)$. The reader can find the specifications of a minimally causal relevant subset and the proof in the work of Malinsky et al. [156]. Equivalence between the structural causal model and the potential outcome framework follows from the equivalent representation of the one-step-ahead counterfactual $V_i(\boldsymbol{pa}_i)$ as the output of the structural equation $f_i(\boldsymbol{pa}_i, \boldsymbol{w}_i)$ (by letting $\boldsymbol{w}_i = \{V_i(\boldsymbol{pa}_i) \mid \boldsymbol{pa}_i \in \Omega_{\mathbf{pa}(V_i)}\}$ and setting $f_i(\boldsymbol{pa}_i, w_i) = (\boldsymbol{w}_i)_{\boldsymbol{pa}_i} = V_i(\boldsymbol{pa}_i)$).

3.5.2 Counterfactual Models

In addition to consistency and causal irrelevance, independence relations are assumed in order to reason about counterfactuals. The literature splits along the lines of which independence assumptions to adopt. There is the more conservative *finest fully randomized causally interpretable structured tree graph* (FFRCISTG) and the more restrictive *non-parametric structural equation model with independent errors* (NPSEM-ie). First, the FFRCISTG independencies are introduced.

Assumption 18 (FFRCISTGS Independencies)**.** Assume one-step-ahead counterfactuals by recursive substitution. Let $\boldsymbol{v}$ be an assignment for random variables $\boldsymbol{V}$ and let $\boldsymbol{pa}_i$ be the restriction of that assignment to parent variables of V_i. Then for each assignment $\boldsymbol{v}$, the corresponding one-step-ahead counterfactuals consistent with $\boldsymbol{v}$ are mutually independent:

$$V_1 \perp\!\!\!\perp_P V_2(\boldsymbol{pa}_2) \perp\!\!\!\perp_P \ldots \perp\!\!\!\perp_P V_n(\boldsymbol{pa}_n),$$

where $V_i < V_{i+1}$ in the topological sort.

It is important to note that all counterfactual random variables are consistent with each other in the sense that there is no contrary assignment among them. Extra independencies across contradicting assignments are imposed by assuming independencies of the error terms in the non-parametric structural equation models. Formally, the counterfactual random variables that are independent in the NPSEM-ie model are defined as follows.

Assumption 19 (NPSEM-ie Independencies)**.** Assume one-step-ahead counterfactuals by recursive substitution. Then the set of one-step-ahead counterfactuals across possibly contradictory interventions are mutually independent:

$$\{V_1\} \perp\!\!\!\perp_P \{V_2(\boldsymbol{pa}_2) \mid \boldsymbol{pa}_2 \in \Omega_{\mathbf{pa}(V_2)}\} \perp\!\!\!\perp_P \ldots \perp\!\!\!\perp_P \{V_n(\boldsymbol{pa}_n) \mid \boldsymbol{pa}_n \in \Omega_{\boldsymbol{pa}(V_n)}\},$$

where $V_i < V_{i+1}$ by the topological sort.

Because the NPSEM-ie independencies also contain the FFRCISTGS independencies, the NPSEM-ie model is *strictly stronger* than the FFRCISTGS model. Consistency and causal irrelevance are implicit in the NPSEM-ie as well as the FFRCISTGS model.

Example 5 (Difference in Counterfactual Model Independencies)**.** Assume one-step-ahead potential outcome random variables corresponding to the nodes $\boldsymbol{Z}, T, Y$ respect-

ing the topological sort of Figure 3.3. Then, following the FFRCISTGS model, for assignment $\boldsymbol{z}_1, t$ the following independence exist:

$$\boldsymbol{Z} \perp\!\!\!\perp_P T(\boldsymbol{z}_1) \perp\!\!\!\perp_P Y(t, \boldsymbol{z}_1).$$

In addition to the previous independencies, according to the NPSEM-ie model, other independencies across contradictory assignments $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ are implied, such as:

$$\boldsymbol{Z} \perp\!\!\!\perp_P T(\boldsymbol{z}_2) \perp\!\!\!\perp_P Y(t, \boldsymbol{z}_1).$$

While DAGs and ADMGs are not expressive enough to account for reasoning with one-step-ahead potential outcomes with either NPSEM-ie independencies or FFRCISTGS independencies, a more refined graphical construction called a *single world intervention graph* (SWIG) was introduced via a node-splitting operation based on causal irrelevance. The SWIG can encode the independence relations of either the NPSEM-ie or the FFRCISTGS. Similarly to how the causal Bayesian networks assume a factorization property of the interventional distributions and modularity property about the nature of interventions, the SWIGs obey properties that specify the behavior of counterfactual distributions. Both the NPSEM-ie model and the FFRCISTGS model, together with consistency, imply these factorization and modularity properties for SWIGs [195] as illustrated by Figure 3.8.

3.5.3 Inference

Inference on the counterfactual level is concerned with the identification of the relevant components necessary to address counterfactual queries. In order to calculate the distribution of counterfactuals under different interventions, the *g-formula* can be used, which has been in Section 3.4.2. This formula can be extended to account for unit-specific interventions and the distribution of that intervention [262] resulting in the *extended g-formula* [198, 195]:

Proposition 20 (Extended G-Formula)**.** Let $\boldsymbol{S} \subset \boldsymbol{V}$ and $V(\boldsymbol{s})$ be the one-step-ahead counterfactual defined by recursive substitution. Then given positivity, the joint distribution can be written as:

$$P(V_1(\boldsymbol{s}), \ldots, V_n(\boldsymbol{s})) = \prod_{i|V_i \in \boldsymbol{V}} P(V_i \mid \boldsymbol{s} \cap \boldsymbol{pa}_i, \mathbf{pa}(V_i) \notin \boldsymbol{S}).$$

The formula is equivalent to the factorization and modularity property implicit in SWIGs and has been proven to hold [195, 221]. The strength of the formula is that it rewrites counterfactual distributions in terms of observational distributions, but unlike the g-formula, the extended g-formula also accounts for nested counterfactuals by having a term for every $V_i \in \boldsymbol{V}$. Analogously, the do-calculus can be extended to rewrite nested counterfactuals such as dynamic treatment regimes or path-specific interventions. For that reason, *po-calculus* has been introduced as a generalization of the do-calculus as a result of consistency, causal irrelevance, and factorization on SWIGs [156]. Although the po-calculus implies the do-calculus for interventional queries [156], it has been shown that additional identification results consisting of nested counterfactuals follow exclusively from the po-calculus [221].

As there exists a hierarchy of causal queries, there is also a *hierarchy of interventions*. The most granular form of interventions are node interventions according to the hierarchy of interventions of [222]. Node interventions are a specific form of edge interventions which in turn are a specific form of path interventions. Multiple targets of interest in mediation analysis are defined as edge interventions and for this reason, the extended g-formula has also been extended to the *edge g-formula* [222]. While node interventions are associated with the FFRCISTG model and require the extended g-formula for identification, edge interventions correspond to the NPSEM-ie model and require the edge g-formula for identification as shown in Figure 3.8.

3.6 Conclusion and Future Work

This section has synthesized existing research on causality by situating different research areas within the framework of Pearl's causal hierarchy. The concepts and associated assumptions required to address queries at different levels of the hierarchy have been highlighted. These foundational causal concepts form the basis for the analyses conducted in the remainder of the thesis. Future research should further explore causal inference in systems with feedback, as most existing work assumes acyclicity in structural causal models, despite real-world phenomena, such as climate systems, often exhibiting cyclical causal relations [55, 36].

Chapter 4

Causal Game Theory

While many causal concepts focus on isolated decision-making, game-theoretic extensions are germane to multi-agent contexts where strategic dynamics are central. These settings often involve actors whose choices depend not only on their own preferences but also on expectations about others' actions. To effectively integrate causal reasoning with game theory in modeling complex security environments, it is essential to understand both the foundational principles of game theory and their intersection with the previously introduced causal concepts.

This chapter systematically maps diverse game forms such as normal form, extensive form, and Bayesian games to their corresponding causal representations, synthesizing previously scattered connections within a probabilistic graphical model framework. Key concepts from (causal) game theory are elucidated, followed by an examination of the input required to operationalize such models. To bridge the gap between theoretical foundations and practical application, the chapter provides structured guidance for practitioners on selecting and applying these models in various security contexts. These concepts are further illustrated with examples derived from complex security environments, particularly those involving the deterrence of adversarial attacks.

In this way, RQ1.3 is addressed: *What methods exist for integrating causal reasoning with strategic decision-making in complex security environments, and how can they be applied?* This analysis not only advances theoretical understanding but also equips practitioners with the tools to model strategic interactions in a causally rigorous manner, fostering more robust decision-making in complex security contexts. Chapter 6 further examines one such tool to illustrate its practical relevance. The content of this chapter closely follows a paper [252].

4.1 Introduction

Game theory examines how rational decision-makers navigate strategic interactions, whether in competitive or cooperative settings. By modeling the decisions of interdependent agents, it provides a structured framework for analyzing incentives, strategy formation, and equilibrium outcomes. Recent research has aimed to combine the strengths of causal modeling and game theory [91, 227].

This chapter provides a structured framework that clarifies key concepts in game theory and its intersection with causality. By consistently adapting a practical example and referencing relevant research for implementation, it offers a concrete guide to navigating the complexities of integrating causal reasoning with game-theoretic modeling. This approach aims to bridge the divide between theory and practice, equipping practitioners with clearer implementation strategies and fostering closer collaboration between researchers and methodologists.

More specifically, the connection between causality and game theory in the context of PGMs [127, 236] is reviewed. The focus is on PGMs as they provide a structured

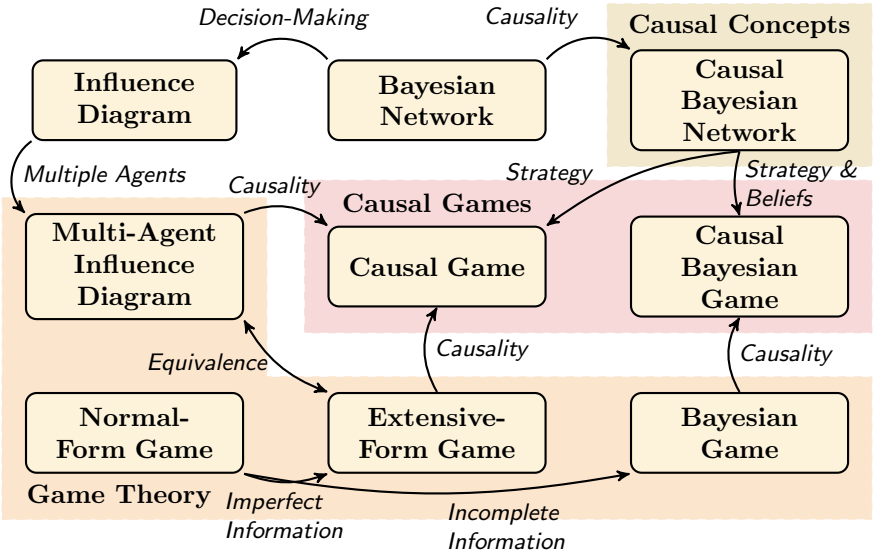


Figure 4.1: Scope of this chapter: the yellow blocks represent key concepts discussed within each domain: causality, game theory, and causal game theory. The concepts are grouped into categories that highlight their primary associations, indicated with background in different colors. Arrows indicate the possible extension or adaptation that allows for the transition from one concept to another.

approach to bridging the gap between theoretical advancements and practical implementation, particularly due to their intuitive representation of dependencies [183] and their capacity to unify diverse cognitive processes through a shared representational framework [58]. Through a detailed examination of their mathematical specifics, supported by illustrative examples from complex security environments, a structured framework that bridges theoretical understanding and practical implementation is provided. While exploring the mathematical details of various game-theoretic models and causal games might seem counterintuitive for practical purposes, it is necessary to articulate the distinctions between these models more precisely. This distinction not only helps practitioners select the most suitable model for their specific application but also clarifies the specific information required to work with these models effectively. Without delving into specifics, research that discusses techniques for eliciting the required information is pointed out. Finally, further considerations and insights are discussed to help surmount practical implementation challenges. The conceptual scope of this chapter is further illustrated in Figure 4.1, which categorizes the key concepts within the (causal) game-theoretical realm and their relation to previously discussed causal concepts.

Although doubts concerning the assumption of agent rationality [214, 83] have led to skepticism regarding the applicability of game-theoretic models [248, 204, 191], this thesis refrains from entering this philosophical debate.

The structure of this chapter is as follows. Section 4.2 introduces game-theoretic models, their solution concepts, and illustrative applications, along with a practical guide to their implementation. Section 4.3 builds on this foundation by extending game-theoretic models into the causal domain, presenting associated solution concepts, and discussing key considerations for their practical application. Conclusions and future research avenues are presented in Section 4.4.

4.2 Game Theory

To align causal concepts with the game-theoretical domain, this section focuses on game-theoretical components that have counterparts in causal reasoning. Therefore, three different types of games are considered: the normal-form game, the extensive-form game, and the Bayesian game. A normal-form game models strategic interactions in which players select actions simultaneously without knowledge of other players' choices, making it well-suited for competitive scenarios such as pricing strategies. An extensive-form game represents sequential decision-making, where players act in a

structured order, as in negotiations. A Bayesian game incorporates uncertainty, allowing players to make decisions based on private beliefs about unknown factors, making it particularly applicable to auctions.

The discussion begins by outlining formal game definitions and relevant solution concepts. Furthermore, the applicability of these game forms is explored by analyzing similar examples across different scenarios. Finally, the challenges associated with each form are addressed, and their practical utility is discussed. An overview of the game forms discussed, their concomitant characteristics, and required information for model implementation are summarized in Figure 4.4.

4.2.1 Normal-Form Game

First, the definition of a normal-form game and its associated solution concept is introduced. Then an example is provided.

Definition 4.1 (Normal-Form Game (NFG))**.** A *normal-form game* is a tuple $\Gamma = (M, \boldsymbol{A}, \boldsymbol{U})$ for which:

- $M = \{1, \ldots, m\}$ is a set of agents.
- $\boldsymbol{A} = \{A^1, \ldots, A^m\}$ is the set of action set, where A^i denotes the set of actions available to agent $i \in M$.
- $\boldsymbol{U} = \{u^1, \ldots, u^m\}$ is a set of utility functions where $u^i : \boldsymbol{A} \to \mathbb{R}$ is the payoff function for agent $i \in M$, representing the payoff that agent i receives.

Definition 4.2 (Nash Equilibrium)**.** A strategy profile $\hat{\sigma} = (\hat{\sigma}^1, \ldots, \hat{\sigma}^m)$ is a *Nash equilibrium* [167] if for every player $i \in \{1, \ldots, m\}$:

$$\hat{\sigma}^i \in \arg\max_{\sigma^i \in \boldsymbol{\Sigma}^i} u^i(\sigma^i, \hat{\sigma}^{-i}).$$

where $\boldsymbol{\Sigma}^i$ is player i's strategy space.

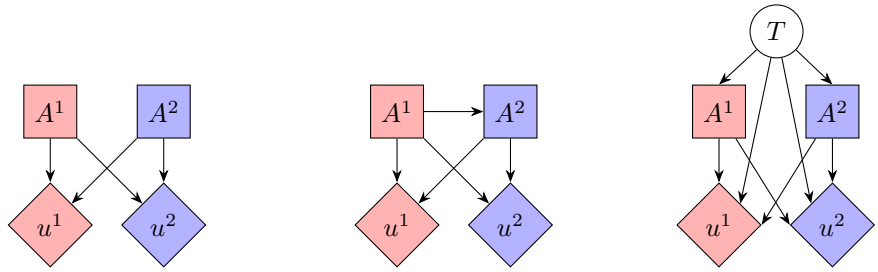
A Nash equilibrium represents a stable state of the game: given that all other players adhere to their equilibrium strategies $\hat{\sigma}^{-i}$, no player i can achieve higher utility by switching to any alternative strategy $\sigma^i \in \boldsymbol{\Sigma}^i$. Each player's equilibrium strategy is thus a best response to the equilibrium strategies of others, and no player has an incentive to unilaterally deviate. An illustration of the Nash equilibrium in a complex security environment follows.

Table 4.1: Utilities in Deterring Game (or Game of Chicken) in Normal-Form

		Adversary	
		a	$\neg a$
Deterrer	d	$(-1000, -1000)$	$(1, -1)$
	$\neg d$	$(-1, 1)$	$(0, 0)$

Example 6 (Deterring Game in Normal-Form)**.** Suppose a game models a strategic interaction between a deterring agent and its attacking adversary. The deterring agent threatens retaliation if attacked, but whether the agent is willing to follow through (d) or is bluffing ($\neg d$) is determined by action set A^1. Simultaneously, the adversary decides whether to attack (a) or not ($\neg a$), denoted by action set A^2. The game can be illustrated by Figure 4.2a and the game matrix is displayed in Table 4.1. The two pure Nash equilibria are ($\neg a$, d) and (a, $\neg d$).

Although the deterring agent acts first by issuing a threat of retaliation, the game-theoretical model does not consider the act of threatening as a strategic decision. Instead, the decision of interest is whether the deterring agent follows through on the threat. Since this decision does not occur after the attacker's move, both actions are effectively simultaneous and can be represented using a normal-form game. Nonetheless, it may very well be that acts do happen sequentially. In this case, a more refined game form is necessary, which is the extensive-form game.



(a) normal-form game **(b)** extensive-form game **(c)** Bayesian game

Figure 4.2: The relations of the normal-form game (a), extensive-form game (b), and Bayesian game (c). These relations correspond to Example 6 for (a), Examples 7 and 8 for (b), and Example 9 for (c). In the normal-form game, the deterring and attacking agents make independent decisions that determine their utilities. In contrast, the attacker's decisions are shaped by the deterrer's actions in the extensive-form game. Lastly, in the Bayesian game, the agents' decisions and utilities are influenced by their individual types and their beliefs about the opponent's type.

4.2.2 Extensive-Form Game

First, the definition of an extensive-form game is introduced, followed by an example and the relevance of a solution concept known as a subgame perfect equilibrium. The definition of Hammond et al. [91] is adopted.

Definition 4.3 (Extensive Form Game (EFG))**.** An *extensive-form game* is a tuple $\Gamma = (M, G, \boldsymbol{P}, \boldsymbol{A}, \lambda, \boldsymbol{I}, \boldsymbol{U})$ for which:

- $M = \{1, \ldots, m\}$ is a set of agents.
- $G = (\boldsymbol{V}, \boldsymbol{E})$ is a rooted tree, where the nodes $\boldsymbol{V}$ are partitioned into sets $\boldsymbol{V^0}, \boldsymbol{V^1}, \ldots, \boldsymbol{V^n}, \boldsymbol{T}$. In this case, $\boldsymbol{T}$ are the leaves or terminal nodes of G, $\boldsymbol{V^0}$ are chance nodes, and $\boldsymbol{V^i}$ are the decision nodes controlled by agent $i \in M$. The nodes are connected by edges $\boldsymbol{E}$.
- $\boldsymbol{P} = \{P_1, \ldots, P_{|\boldsymbol{V}^0|}\}$ represents a set of probability distributions P_j defined over the children of each chance node V_j^0, denoted as $\mathbf{ch}(V_j^0)$, for $j = 1, 2, \ldots, |\boldsymbol{V}^0|$.
- $\boldsymbol{A}$ represents the set of action sets, where $A_j^i \subseteq \boldsymbol{A}$ indicates the set of actions available at $V_j^i \in \boldsymbol{V^i}$.
- $\lambda : \boldsymbol{E} \to \boldsymbol{A}$ is a labelling function that assigns each edge (V_j^i, V_l^k) to an action $a \in A_j^i$.
- $\boldsymbol{I} = \{\boldsymbol{I}^1, \ldots, \boldsymbol{I}^m\}$ represents a collection of information sets, which partition the decision nodes controlled by agent i. Each information set $I_j^i \in \boldsymbol{I}^i$ is defined such that, for all $V_k^i, V_l^i \in I_j^i$, the available actions at these nodes are identical, i.e., $A_k^i = A_l^i$.
- $\boldsymbol{U} = \{u^1, \ldots, u^m\}$ is a set of utility functions where $u^i : \boldsymbol{T} \to \mathbb{R}$ is the payoff function for agent $i \in M$, representing the payoff that agent i receives.

Example 7 (Deterring Game in Extensive-Form)**.** Suppose a game models a strategic interaction between a deterring agent and its attacking adversary. This time, the deterring agent aims to deter by threatening retaliation denoted in action set A^1 in node V_1^1, which it is willing to follow through (d) or is bluffing ($\neg d$). Subsequentially, the adversary then acts A^2 to decide whether to attack (a) or not ($\neg a$) in nodes (V_1^2, V_2^2). Since the adversary does not have knowledge about the credibility of the deterrence effort, both nodes (V_1^2, V_2^2) are in the same information set, I_1^2. The dependencies of the game can be illustrated by Figure 4.2b and the game tree of Figure 4.3.

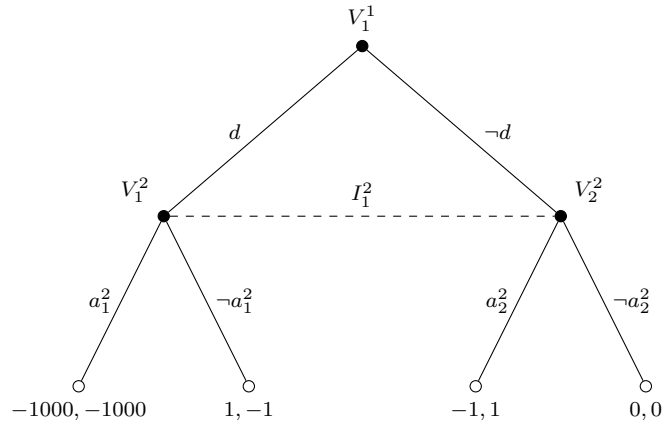


Figure 4.3: The figure illustrates a game tree of the deterring game in extensive-form.

Should the adversary have full information about the action of the deterring agent, a new game arises in which the adversary evaluates its actions based on this knowledge. Formally, this is known as a *subgame*.

Definition 4.4 (Subgame)**.** A *subgame* of an EFG is a game with a game tree $G' = (\boldsymbol{V}', \boldsymbol{E}')$ that is restricted to a node and its descendants such that any information set is either completely in the subgame or completely out the subgame.

Definition 4.5 (Subgame Perfect Equilibrium (SPE))**.** A *subgame perfect equilibrium* is a strategy profile σ such that for every subgame, σ constitutes a Nash equilibrium of the subgame.

Subgame perfect equilibria are relevant for the exclusion of non-credible threats, which are threats that a rational player has no incentive to carry out in later stages of the game. This is illustrated by the following example:

Example 8 (Deterring Subgame in Extensive-Form)**.** Consider a slightly modified version of Example 7 under conditions of perfect information, where the adversary recognizes the deterrer's commitment to enforce the threat, as shown in Figure 4.2b. The Nash equilibria are $((\neg a_1^2, \neg a_2^2), d), ((\neg a_1^2, a_2^2), d)$ and $((a_1^2, a_2^2), \neg d)$. In the subgame, the deterring agent anticipates that the adversary's threat to attack is not credible in the case of committed punishment, leaving $((\neg a_1^2, a_2^2), d)$ as the only subgame perfect equilibrium.

While *imperfect information* (uncertainty about the history of play) is captured directly through information sets, *incomplete information* (uncertainty about game

structure) requires modeling player types and beliefs. Bayesian games provide the framework for this.

4.2.3 Bayesian Game

The formal definition of a Bayesian game is first presented, followed by the introduction of its solution concept, the Bayesian Nash equilibrium. An example is then provided that extends the previous illustrations.

Although Bayesian games are sometimes introduced in terms of nature's states [174], the formulation in terms of players' types, as originally proposed [92, 76], is adopted here.

Definition 4.6 (Bayesian Game)**.** A *Bayesian game* is a tuple $\Gamma = \{M, \boldsymbol{A}, T, P, \boldsymbol{U}\}$, such that:

- $M = \{1, \ldots, m\}$ is the finite player set.
- $\boldsymbol{A}$ represents the set of action sets, where A^i is the action set of player i for $i \in M$.
- T^i is the finite type set of player i, and $t^i \in T^i$ its type. $T = (T^1, \ldots, T^m)$ is called the type profile tuple of Γ.
- $P : T \to [0, 1]$ is a probability distribution over T, referred to as the common prior. The belief of player i is denoted by

$$P(t^{-i} \mid t^i) = \frac{P(t^{-i}, t^i)}{P(t^i)} = \frac{P(t^{-i}, t^i)}{\sum_{t^{-i}} P(t^{-i}, t^i)},$$

  which describes player i's uncertainty about the other $m - 1$ players' possible types t^{-i}, given player i's type t^i, where $t^{-i} = (t^1, \ldots, t^{i-1}, t^{i+1}, \ldots, t^m)$ represents the tuple of the types of all the players except for player i.
- $\boldsymbol{U} = \{u^1, \ldots, u^m\}$ is a set of utility functions, where $u^i : T \times \boldsymbol{A} \to \mathbb{R}$ is the payoff function, which maps each action profile $\boldsymbol{a} \in \boldsymbol{A}$ to the pay-off of player i under each type profile $t^i \in T^i$.

Now that the structure of the game is contingent on the types of the players, the concept of a behavioral strategy can naturally be extended to account for these types: $\sigma^i(t^i) := \sigma^i(A^i \mid t^i)$ [22]. This allows the definition of a *Bayesian Nash equilibrium*.

Table 4.2: Utilities in Deterring Game in Bayesian Game

		Adversary Type t^1: $p = \frac{1}{4}$		Adversary Type t^2: $p = \frac{3}{4}$	
		a	$\neg a$	a	$\neg a$
Deterrer	d	$(-1000, -1000)$	$(1, -1)$	$(-1, 0)$	$(1, -1)$
	$\neg d$	$(-1, 1)$	$(0, 0)$	$(-1, 1)$	$(0, 0)$

Definition 4.7 (Bayesian Nash Equilibrium)**.** A strategy profile $\hat{\sigma} = (\hat{\sigma}^1, \ldots, \hat{\sigma}^m)$ is a *Bayesian Nash equilibrium* if for every player $i \in M$ and type $t^i \in T^i$:

$$\hat{\sigma}^i(t^i) \in \arg\max_{\sigma^i \in \mathbf{\Sigma}^i} P(t^{-i} \mid t^i) u^i(t^i, t^{-i}, \sigma^i, \hat{\sigma}^{-i}(t^{-i})).$$

Example 9 (Deterring Game in Bayesian Form)**.** Similar to the normal-form game, the actions of the deterring agent A^1 and its adversary A^2 are modeled independently. This time, the adversary can assume different types, where it is either protected against retaliation (t^2) or not (t^1). The relations of the game are illustrated by Figure 4.2c and the game matrices displayed in Table 4.2. The Bayesian Nash equilibria are $((\neg a, a), d)$ and $((a, a), \neg d)$. These are found by verifying that the deterrer maximizes expected utility given the probability distribution over adversary types ($p(t^1) = \frac{1}{4}$, $p(t^2) = \frac{3}{4}$), while each adversary type simultaneously maximizes their own payoff given their private information and the deterrer's strategy.

4.2.4 Practical Guide to Game Theory

Game theory offers a structured framework for organizing information on actors' decision-making processes [99]. Before applying game-theoretic concepts, a qualitative process should define the specific rules of the game for a given problem. This involves identifying stakeholders, outlining potential policy options, and establishing their interdependencies. Such an approach to framing policy problems is known as *metagame analysis* [107].

Practitioners should select the game type that best fits the policy problem's structure. Simultaneous decisions suit normal-form and Bayesian games,[1] whereas sequential decisions align with extensive-form games. Additionally, extensive-form games may involve imperfect information, while Bayesian games feature incomplete information. Although selecting a specific game type is useful, some scholars argue that a robust analytical approach benefits from modeling pluralism rather than strict unifor-

[1] Bayesian games can also be extended to sequential decision-making [174].

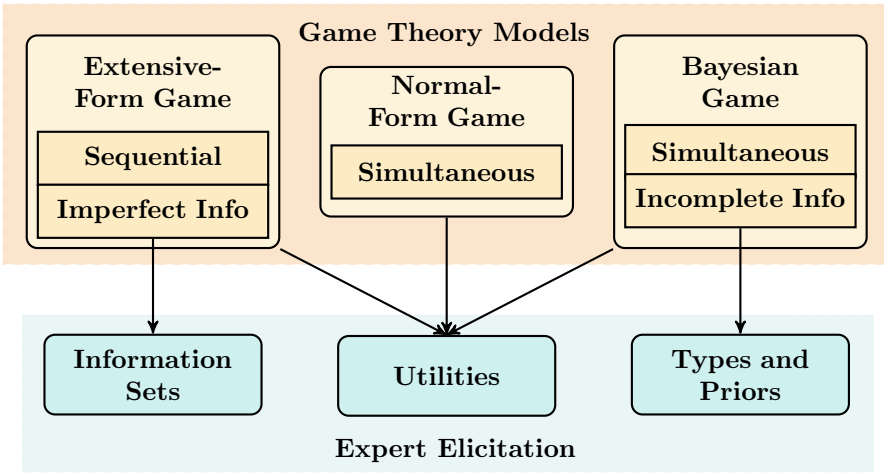


Figure 4.4: Elicitation requirements for different game-theoretic models: while the yellow blocks correspond to the different games along with their associated properties, the light green blocks indicate the different pieces of information that are required to be elicited. The arrows indicate what information is relevant to each game.

mity [6], advocating for the use of diverse game-theoretic frameworks where necessary while ensuring that variation remains focused on essential structural distinctions.

As the mathematical dissection of different games shows, each game requires specific information to be gathered before implementation. All games need the elicitation of utilities or preferences, with extensive-form and Bayesian games requiring more detailed utility structures. While utility elicitation is a demanding task, it has been well-studied and the reader is referred to the work by Wakker and Deneffe [254] for details about elicitation methods. If preference relations are uncertain, alternative methods can still provide insights into the stability of certain solution concepts [143].

In extensive-form games, information sets—clarifying who knows what at each decision point—need to be extracted. Additionally, chance nodes require probability distributions for the uncertainties they introduce. For Bayesian games, prior and posterior probabilities of types must also be elicited, a challenging process with guidelines written by Mikkola et al. [161]. The characteristics of the different games, along with the required elicitation information can be observed in Figure 4.4.

To bridge the gap between theory and application in game theory, a growing body of work focuses on deriving game-theoretic models from simulation data to enhance empirical validity. A comprehensive survey of recent advances in this area is provided by Wellman et al. [257].

4.3 Causal Game Theory

In this section, the intersection of game theory and causality is explored by integrating causal concepts from the previous chapter into the decision-making framework. Specifically, the Bayesian networks are extended to influence diagrams [108], distinguishing between purely probabilistic structures and decision-theoretic elements. This framework is further generalized to multi-agent settings, leading to multi-agent influence diagrams [128, 91], which form the foundational structure of causal games. To account for uncertainty in the causal structure of a game, the approach of Gonzalez Soto et al. [227] is adapted, introducing the causal Bayesian game. Finally, key considerations for the practical implementation of these models are discussed.

To establish a foundation for causal game theory, previous examples are reformulated through a causal lens. This provides a basis for extending causal reasoning into the strategic reasoning domain.

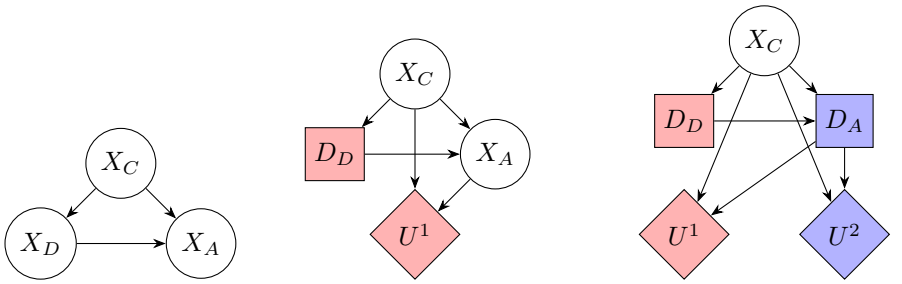
Example 10 (Deterring Relations in Causal Form)**.** Suppose there is observational data on the explicitness of deterrence messages X_D, which can either be explicit (d) or vague ($\neg d$). The goal of an explicit message is to dissuade the adversary from committing to an aggressive operation X_A ($X_D \to X_A$). However, both the explicitness of the deterrence message and the adversary's decision to attack are shaped by the deterring agent's military and strategic capabilities X_C, which can be strong (c) or weak ($\neg c$). These capabilities influence the explicitness of the message $X_C \to X_D$, but also directly affect the adversary's decision to commit to an aggressive operation $X_C \to X_A$. The causal relations are displayed in Figure 4.5a.

To compute the causal effect of explicit deterrence messaging on successful dissuasion, the covariate, the deterrer's capabilities, should be adjusted for. Therefore, the *do*-operator can be deployed, and the truncated factorization can be utilized:

$$P(X_A \mid do(X_D = d)) = \sum_{x_C \in \{c, \neg c\}} P(X_A \mid X_D = d, X_C = x_C) P(X_C = x_C).$$

4.3.1 Influence Diagram

An influence diagram extends a Bayesian network to the decision-making realm by dissecting the nodes into chance nodes, utility nodes, and decision nodes. More formally:



(a) Bayesian Network **(b)** Influence Diagram **(c)** MAID

Figure 4.5: The relations of the Bayesian network (a), influence diagram (b), and multi-agent influence diagram (c) of Example 10, 11 and 12, respectively. The Bayesian network does not distinguish between chance, decision and utility nodes as do the influence diagrams. In addition, the multi-agent influence diagram models the decision of the adversary also strategically.

Definition 4.8 (Influence Diagram)**.** An *influence diagram* contains a graphical structure $G = (\boldsymbol{V}, \boldsymbol{E})$ where $\boldsymbol{V}$ is separated into decision nodes $\boldsymbol{D}$, chance nodes $\boldsymbol{X}$ and utility nodes $\boldsymbol{U}$. Whereas the conditional probability distributions of $\boldsymbol{X}$ and $\boldsymbol{U}$ are known, any *decision rule* $\sigma(D)$ with $D \in \boldsymbol{D}$ corresponds to a conditional probability distribution over the decisions and hence all decision rules constitute the full joint probability distribution $P(\boldsymbol{V})$.

Example 11 (Deterring Relations as Influence Diagram)**.** A refinement of Example 10 involves separating the chance node of the military and strategic capabilities X_C from the decision node of deterrence messaging D_C. In this framework, the aggressor's decision to conduct an aggressive operation X_A can be modeled as another chance node, which is followed by a final utility node of the deterring agent U^1. These relations are illustrated by Figure 4.5b.

Influence diagrams incorporate decision-making elements into Bayesian networks,[2] but they only model the decision-making of a single agent, meaning there is no strategic interaction between agents. Strategic considerations emerge only when multiple agents make decisions in response to each other, as seen in multi-agent influence diagrams.

4.3.2 Multi-Agent Influence Diagram and Causal Game

First, multi-agent influence diagrams [128] are introduced, followed by the definition of a causal game [91]. The latter concept will then be illustrated with an example.

[2] In a similar way, causal Bayesian networks can be extended to *causal influence diagrams* [75].

Definition 4.9 (Multi-Agent Influence Diagram (MAID))**.** A *multi-agent influence diagram* contains a graphical structure $G = (\boldsymbol{V}, \boldsymbol{E})$ and a set of agents $M = \{1, \ldots, m\}$. Furthermore, the nodes $\boldsymbol{V}$ are separated in decision nodes $\boldsymbol{D} = \cup_{i \in M} \boldsymbol{D}^i$, chance nodes $\boldsymbol{X}$ and utility nodes $\boldsymbol{U} = \cup_{i \in M} \boldsymbol{U}^i$. Each strategy $\sigma^i(D^i)$ with $D^i \in \boldsymbol{D^i}$ defines a conditional probability distribution over a decision node. Consequently, given conditional probability distributions of $\boldsymbol{X}$ and $\boldsymbol{U}$, a complete strategy profile σ constitutes the full joint probability distribution $P(\boldsymbol{V})$.[3]

The causal game Γ associated with a MAID can be seen as a more abstract form of a MAID, where the parameters of the decision variables are yet to be defined.

Definition 4.10 (Causal Game)**.** A *causal game* Γ is a MAID such that for any chosen strategy profile σ, the induced joint probability distribution $P^\sigma(\boldsymbol{V})$ corresponds to a causal Bayesian network.

Similarly to extensive-form games, the Nash equilibrium of causal games can be defined in terms of the strategy profiles:

Definition 4.11 (Nash Equilibrium)**.** A strategy profile $\hat{\sigma} = (\hat{\sigma}^1, \ldots, \hat{\sigma}^m)$ is a *Nash equilibrium* if for every player $i \in \{1, \ldots, m\}$:

$$\hat{\sigma}^i \in \arg\max_{\sigma^i \in \boldsymbol{\Sigma}^i} \sum_{U \in \boldsymbol{U}^i} \mathbb{E}_{[\sigma^i, \hat{\sigma}^{-i}]}[U].$$

Multi-agent influence diagrams are powerful models, because they allow for the calculation of equilibria as well as the computation of policy interventions.

Example 12 (Multi-Agent Influence Diagram Deterring Game)**.** A multi-agent influence diagram can be derived from Example 11 when the adversary's decision to conduct an aggressive operation a or refrain from one $\neg a$ is modeled as a decision node of another agent D_A. This decision node is influenced by the deterring agent's decision node, D_D, which represents whether the deterrence messages are explicit (d) or vague ($\neg d$). Alongside the deterring agent's utility node U^1, the adversary also

Figure 4.6: Utilities of the Deterring Agent (left) and Attacking Agent (right)

U^1	$X_C = c$	$X_C = \neg c$
$D_A = a$	0	-1
$D_A = \neg a$	1	1

U^2	$X_C = c$	$X_C = \neg c$
$D_A = a$	-1000	1
$D_A = \neg a$	-1	-1

[3]Since EFGs and MAIDs are proven to be equivalent [91], σ is used again for a strategy profile.

possesses a utility node U^2. Both utility nodes depend on the attacker's decision D_A and the deterrer's capabilities X_C, which can be either strong (c) or weak ($\neg c$), with an equal probability distribution. The relations are displayed in Figure 4.5c and the utilities are further specified in Figure 4.6.

The game has eight Nash equilibria, one of which is equilibrium $\hat{\sigma}$ where the deterring agent issues an explicit deterrence message when possessing strong capabilities and a vague one otherwise. In this scenario, the adversary chooses not to attack if and only if the deterring agent demonstrates strong capabilities regardless of the deterrence message. The equilibrium $\hat{\sigma}$ also happens to be a subgame perfect equilibrium.[4] Within this equilibrium, the expected utility of the deterring agent for sending out an explicit message is $\mathbb{E}_{[\hat{\sigma}]}[U^1 \mid D_D = d] = 1$. Intervention effects can also be assessed in this equilibrium; for instance, if allied agents force an explicit deterrence message regardless of its capability, the utility becomes $\mathbb{E}_{[\hat{\sigma}]}[U^1 \mid do(D_D = d)] = 0$.

This example is a *post-policy* intervention as the results are computed after a strategy profile from the Nash equilibrium has been chosen. In contrast, *pre-policy* interventions allow agents to adjust their strategy profile after an intervention, which requires a more refined notion of a MAID [91]. The introduction of these notions, while relevant for strategic reasoning, is considered beyond the scope of the chapter as they do not bring additional implications for their practical implementation.

4.3.3 Causal Bayesian Games

The notion of a causal Bayesian game [227] was developed in order to allow for uncertainty about a graphical structure controlling an environment in which agents are located. The definition and notation of Soto et al. [227] is refined to align with the previously introduced causal games while ensuring consistency with earlier introduced Bayesian games. Following the Bayesian game in Section 4.2.3, the types correspond to distinct MAIDs within the family of causal graphical structures $\mathcal{G}$. Moreover, as in Bayesian games, players act independently, implying that a subset of MAIDs is considered where no direct causal paths exist between the decision nodes of different agents. Unlike the earlier introduced Bayesian game, players agree on the common state space $\mathcal{G}$ of possible graphical models but have private beliefs about the probability of these states $\mu_i(\mathcal{G})$. Naturally, interventions on decision nodes induce variations in payoffs across different graphical models $G \in \mathcal{G}$.

[4] Although the subgame perfect equilibrium of Definition 4.5 naturally extends to causal games, the notion of subgames in causal games is much richer than in EFGs. The reader is referred to the work of Hammond et al. [91] for details on this.

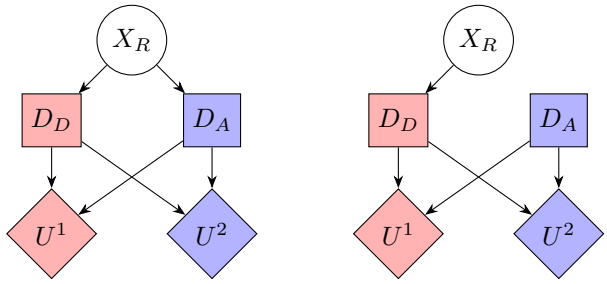
Definition 4.12 (Causal Bayesian Game)**.** Consider a family of different causal structures $G \in \mathcal{G}$ where no direct paths exist between the decision nodes of different agents. Each agent $i \in \{1, \dots, m\}$ has a private belief about the probability of these causal structures $\mu_i(\mathcal{G})$ and a higher-order belief $\mu_i(\mu_{-i}(\mathcal{G}))$, which reflect uncertainty over other $m-1$ players' beliefs. Each strategy $\sigma^i(G) = \sigma^i(D^i \mid G)$ defines a conditional probability distribution over a decision node $D^i \in \boldsymbol{D^i}$ conditional on the belief $\mu_i(G)$ of that graphical model $G \in \mathcal{G}$ for agent $i \in \{1, \dots, m\}$.[5] A *causal Bayesian game* Γ is a MAID such that for any chosen strategy profile σ, the induced joint probability distribution $P^{\sigma}(\boldsymbol{V})$ corresponds to a causal Bayesian network.

Note that the causal Bayesian network is not only induced by the strategies for each player but by the strategies of the players conditioned on the same graphical model. Naturally, these considerations are also reflected in the Bayesian Nash equilibrium.

Definition 4.13 (Bayesian Nash Equilibrium)**.** A strategy profile $\hat{\sigma} = (\hat{\sigma}^1, \dots, \hat{\sigma}^m)$ is a *Bayesian Nash equilibrium* if for every player $i \in \{1, \dots, m\}$ and graphical structure $G \in \mathcal{G}$:

$$\hat{\sigma}^i(G) \in \arg\max_{\sigma^i \in \boldsymbol{\Sigma}^i} \mu_i(\mu_{-i}(G)) \sum_{U \in \boldsymbol{U}^i} \mathbb{E}_{[\sigma^i(G), \hat{\sigma}^{-i}(G)]}[U].$$

Example 13 (Causal Bayesian Deterring Game)**.** Suppose two agents have their private beliefs about the causal structure of the game they are playing. In both games, players make decisions to defend D_D and attack D_A independently. While the utility of



(a) MAID G_1 **(b)** MAID G_2

Figure 4.7: The different causal structures G_1 and G_2 of the causal Bayesian game as in Example 13.

[5] Although the *do*-operator is applied in the original paper, it is relaxed here to allow for the implementation of more dynamic strategies.

Figure 4.8: Utilities of the Deterring Agent (left) and Attacking Agent (right) for Causal Structure G_1 (top) and G_2 (bottom)

$U^1(G_1)$	$D_A = a$	$D_A = \neg a$
$D_D = d$	-1000	1
$D_D = \neg d$	-1	0

$U^2(G_1)$	$D_A = a$	$D_A = \neg a$
$D_D = d$	-1000	-1
$D_D = \neg d$	1	0

$U^1(G_2)$	$D_A = a$	$D_A = \neg a$
$D_D = d$	-1	1
$D_D = \neg d$	-1	0

$U^2(G_2)$	$D_A = a$	$D_A = \neg a$
$D_D = d$	0	-1
$D_D = \neg d$	1	0

both agents is the result of both agents' actions, the attacking agent's decision can also be shaped by the defending agent's capability to retaliate X_R. Figure 4.7 illustrates the two different causal structures under consideration. While the attacking agent does not have access to the defending agent's retaliation capacity nor does he think the defending agent thinks he has ($\mu_A(G_2) = 1$ and $\mu_A(\mu_D(G_1)) = 1$), the defending agent considers a scenario where the attacking has access to his retaliation capacity with equal probability: $\mu_D(G_1) = \mu_D(G_2) = \mu_D(\mu_A(G_1)) = \mu_D(\mu_A(G_2)) = \frac{1}{2}$. Taking into account the utilities for the different structures indicated by Table 4.8, the Bayesian Nash equilibria are $((\neg a,a), d)$ and $((a,a), \neg d)$.

While this game and associated Bayesian Nash equilibrium is similar to Example 9, it is important to emphasize that uncertainty about the causal structure in this example only gives rise to different pay-offs. When alternative causal structures yield distinct payoff configurations and more sophisticated higher-order beliefs are involved, significant complications may arise.

4.3.4 Practical Guide to Causal Game Theory

Analogous to selecting a game-theoretic model, the choice of a causal game-theoretic model should consider whether agents possess private information regarding the causal structure. As the mathematical dissection discerns different types of nodes within causal games, practitioners must also clearly differentiate between decisions, chance events, and utilities. This distinction can be subtle, as illustrated in Example 12: a deterrer's capability, often modeled as a chance node, may not truly qualify as such if the agent has the option to enhance their capabilities. Consequently, this classification requires careful consideration, thoughtfully aligned with the specific research question at hand.

Unlike the standard Bayesian network in Example 10, which consists solely of

chance nodes, causal games with decision nodes do not require the elicitation of conditional probability distributions for those decisions, as they are being solved in response to the adversary. This game-theoretic aspect in causal games thus reduces some of the elicitation burden. The remaining conditional distribution of the chance nodes and the specifications in the utility nodes can be extracted via the elicitation methods introduced in Section 3.4.4 and Section 4.2.4, respectively.

While extracting higher-order beliefs in addition to uncertainty over the nature of graphical structure may appear highly complex, the methods for eliciting prior and posterior probabilities outlined in Section 4.2.4 remain applicable [161]. However, the increased complexity associated with calculating the relevant solution concepts across different causal structures may impede practical implementation.

A summary of the causal game and the causal Bayesian game along with required information for implementation is given in Figure 4.9.

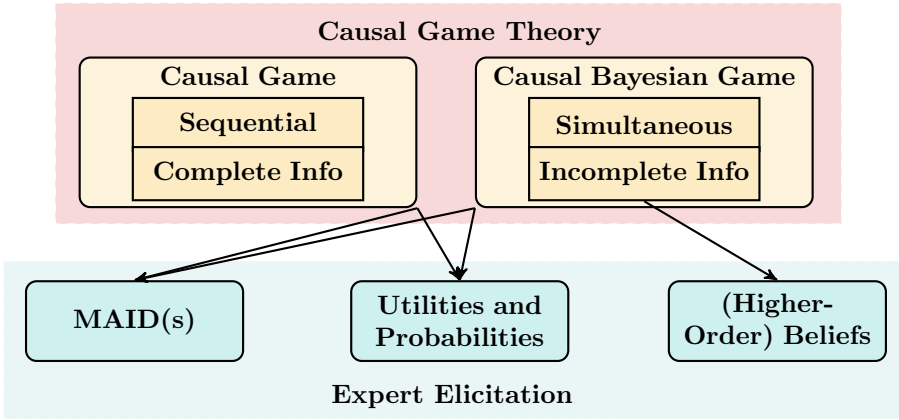


Figure 4.9: Elicitation requirements for causal games and causal Bayesian games and associated characteristics: a causal game necessitates the elicitation of a MAID, utilities, and conditional probabilities. In contrast, a causal Bayesian game further requires the elicitation of multiple MAIDs along with beliefs about the probabilities over these graphs and higher-order beliefs about other players' beliefs.

4.4 Conclusion and Future Work

This chapter has examined game-theoretical models and their integration with previously introduced causal concepts within the framework of probabilistic graphical models. The distinctions between various model types and the input required for their

implementation have been clarified. Practical guidelines, detailed examples, and considerations for effective application have also been provided. Future research could focus on developing integrated tools that unify model elicitation and implementation, streamlining the application of causal game-theoretical models [120, 29].

Chapter 5

Optimization of Causal Interventions

Beyond establishing the methodological foundations of causality and game theory, this dissertation aims to equip practitioners with the necessary tools to engage with key aspects of causal reasoning. In particular, the focus is on the optimization of causal interventions, examining how they can be computed efficiently while maintaining high accuracy.

As associational queries can already be computationally demanding, causal interventions often require the computation of multiple associational queries and are therefore even more demanding. Optimizing over multiple causal interventions constitutes a particularly burdensome computational task.

This chapter provides a computationally efficient methodology to optimize over multiple causal interventions in Markovian Bayesian networks. This chapter thereby addresses RQ2: *How can optimal causal interventions be computed with high accuracy while ensuring computational efficiency?* The content of this chapter closely aligns with two peer-reviewed conference papers [249, 253].

5.1 Introduction

Before presenting the approach to optimizing causal interventions, the problem is first formally defined in terms of the concepts introduced in Chapter 3. Additionally, existing methodologies are reviewed for addressing this problem, highlighting their

strengths and limitations.

Similar to how influence diagrams make a separation between decision, utility and chance nodes, the variables in a causal Bayesian network $\boldsymbol{V}$ can be further dvidided into intervenable variables $\boldsymbol{D}$, context variables $\boldsymbol{X}$, and outcome variable Y.[1] The goal is then to optimize the interventions among intervenable variables that minimize the expected value of the outcome variable in the associated interventional distribution:

$$\boldsymbol{V}_K^*, \boldsymbol{v}_K^* = \underset{\boldsymbol{V}_K \subset \boldsymbol{D}, \boldsymbol{v}_K \in \Omega_{\boldsymbol{V}_K}}{\arg\min} \mathbb{E}[Y \mid do(\boldsymbol{V}_K = \boldsymbol{v}_K))]. \tag{5.1}$$

This is called the offline *causal global optimization* problem [4] on Markovian Bayesian networks with continuous as well as discrete variables. The problem can be separated into two parts. First, there is a necessity to compute the inference queries $\mathbb{E}[Y \mid do(\boldsymbol{V_K} = \boldsymbol{v_K}))]$ efficiently in terms of accuracy and in terms of computational cost. Second, this computation procedure should then be embedded in an optimization framework that can formulate an answer to the causal global optimization problem.

Computing interventional queries requires applying the adjustment formula from Chapter 3. When the set of adjusted variables $\boldsymbol{V}_J \subset \boldsymbol{V}$ is non-empty and continuous, the interventional distribution is given by:

$$p(Y \mid do(\boldsymbol{V}_K = \boldsymbol{v}_K)) = \int_{\Omega_{\boldsymbol{V}_J}} p(Y \mid \boldsymbol{v}_K, \boldsymbol{v}_J) p(\boldsymbol{v}_J) d\boldsymbol{v}_J. \tag{5.2}$$

Even though many Bayesian network applications require the accommodation of such continuous variables [65, 162], state-of-the-art methods for continuous or hybrid (combination of discrete and continuous) Bayesian network inference are still underdeveloped. Algorithms have been developed to conduct inference on hybrid Bayesian networks when a conditional Gaussian distribution among the variables is assumed [127]. However, assuming the parametric form of the distribution is costly, which is why much research has been dedicated to approximation by either discretizing Bayesian networks [31, 169, 168] or by approximating the distribution of the variables in the Bayesian network with a linear combination of exponentials [206] or polynomials [216], which both allow inference.

Discretization of the continuous variables enables the use of established discrete Bayesian network inference methods. Variable elimination and belief propagation are

[1] While this classification closely resembles the structure of influence diagrams, the framework considered here is technically a causal Bayesian network. This distinction arises from the assumption that decision rules are fixed, meaning the decision-maker is limited to performing hard interventions on variables rather than selecting decision rules.

well-developed exact inference methods for discrete Bayesian networks that exploit the structure of the Bayesian network to substantially reduce the computational burden. Nevertheless, even with these effective algorithms, the computational cost increases exponentially as the number of parent nodes within the network grows. Therefore, researchers often employ approximate methods such as sampling or variational inference approaches for more complex Bayesian networks. These methods are summarized by Koller and Friedman [127].

While discretization allows the use of discrete Bayesian network inference algorithms, it may lead to a loss of information, resulting in a lower accuracy of the inference query. At the same time, the computational cost of inference depends heavily on the number and positioning of bins that result from the discretization process. Or, as stated in Koller and Friedman [127], "discretization provides a trade-off between the accuracy of the approximation and cost of computation."

To address the computational challenges of Bayesian network inference after discretization, knowledge compilation [59] can be used. In knowledge compilation, information (such as the probability distribution given by a Bayesian network) is translated without loss into a format that can be queried efficiently. One of the motivations behind knowledge compilation is that by first performing a potentially computationally expensive 'compilation' step, which takes exponential time in the worst case, afterwards the result of many queries (such as inference queries) can be computed quickly. While compilation of the Bayesian network is a heuristic compression method, guided by practical effectiveness rather than formal guarantees of reduced inference complexity, it often proves much faster in practice. This property is particularly advantageous when inference is embedded within an optimization framework, where numerous inference queries must be evaluated efficiently. Specifically, the compilation of Bayesian networks into *binary decision diagrams* (BDDs) [40] is considered, as they have been shown to perform well in the context of Bayesian network inference [56].

The entire methodology is first discussed in Section 5.2, including discretization, knowledge compilation methods, and optimization. As discretization is associated with a loss of information, experiments are run that provide insight into the trade-off between the accuracy of the inference approximation/discretization and the cost of its computation. Further experiments were conducted embedding this framework within various optimization heuristics, in order to evaluate the performance of different algorithms when optimizing over causal interventions. The experimental setup, including the evaluation metrics and considered Bayesian networks for both experiments, is outlined in Section 5.3. The results are presented in Section 5.4, followed by concluding

remarks in Section 5.5.

5.2 Methodology

In this section, a methodology is proposed to tackle the offline causal global optimization problem on hybrid Bayesian networks. The first step is to encode discretized versions of the Bayesian networks as binary decision diagrams. These binary decision diagrams are subsequently subjected to heuristic optimization algorithms that use these efficient encodings to optimize over interventional queries.

More specifically, the entire methodology, including evaluation metrics, is specified in Figure 5.1.[2] To allow a different number and positioning of discretized bins, different types of discretization methods are considered: two unsupervised approaches,

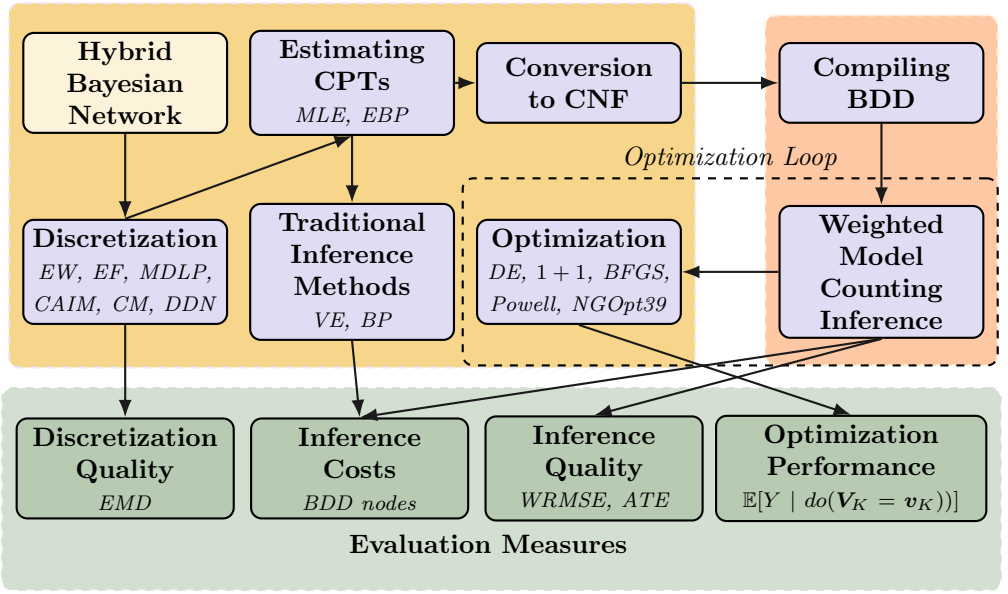


Figure 5.1: Methodology for offline causal global optimization on hybrid Bayesian networks. Python-based components (yellow) handle discretization, estimating CPTs, traditional inference, conversion to CNF formulas, and optimization. C-based components (orange) compile and query BDDs for efficient interventional inference. Evaluation measures (green) assess discretization quality, inference cost, inference accuracy, and optimization performance.

[2]Implementation of the methodology is available on `https://github.com/sebastiaanbrand/bn-dd`.

equal width (EW) and *equal frequency* (EF) binning, as well as the supervised *minimum description length principle* (MDLP) binning method, *class-attribute interdependence maximization* (CAIM) discretization [133], *ChiMerge* (CM) discretization [124] and *dynamic discretization* (DDN) [168]. Conditional probability tables of discretized Bayesian networks are inferred via *maximum likelihood estimation* (MLE) as well as via a *Bayesian method with adjusted empirical Bayes priors* (EBP). Subsequently, the discretized Bayesian networks are encoded as *conjunctive normal form* (CNF) formulas, which are then compiled into BDDs. Inference with BDDs first requires performing a computationally expensive compilation step, after which computing inference queries can be done in time linear in the size of the BDD [57]. As an empirical demonstration, a range of black-box optimization algorithms from the literature [19] is applied,[3] including *random search* as a baseline, *differential evolution* [235] (DE), *(1+1) evolutionary strategy* [32], *BFGS* [39], *Powell's method* [186] and the *NGOpt39 optimizer* [160].

In applying the methodology to a variety of hybrid Bayesian networks, several findings are reported. First, the computation time of BDD-based inference is compared with traditional inference methods such as *variable elimination* (VP) and *belief propagation* (BP). Second, the trade-offs between the quality of the discretization/inference and the computational cost of BDD-based inference algorithms is studied. Finally, the performance of several heuristic optimization algorithms in addressing the offline causal global optimization problem within this BDD-based inference framework is reported.

To assess the trade-off between the quality of discretization/inference and the cost of knowledge compilation, a concept known in multi-objective optimization as the *Pareto front* is used to visualize the results of various considered approaches. A Pareto front represents the set of non-dominated solutions where improving one objective would result in degrading another. The evaluation involves measuring discretization quality in terms of the *earth mover's distance* (EMD) and quantifying knowledge compilation cost by considering the number of nodes in the BDD. Additionally, for non-causal networks, the quality of conditional queries (if ground truth is available) is measured using the *weighted root mean squared error* (WRMSE). For causal Bayesian networks, such additional quality evaluation is done via the *percentage error* of the average treatment effect (PE ATE). Finally, the optimization is evaluated by the objective function (interventional query) with respect to the number of evaluations.

[3]The experiments use implementations from the Nevergrad library, an open-source platform for optimization available at `https://github.com/facebookresearch/nevergrad`.

While the datasets to which the methodology is applied, along with the evaluation measures, are further discussed in Section 5.3, the discretization, parameter learning, BDD encoding, and optimization process are discussed in more detail in this section. A general limitation in this area is the lack of benchmark datasets with known ground truths that follow general continuous distributions. To address this, we use synthetic data generated from continuous distributions with known ground truths, as well as observational data with corresponding experimental counterparts, both of which enable meaningful validation.

5.2.1 Discretization and Parameter Learning Methods

The discretization process serves to partition the state space Ω_{X_i} of a continuous random variable X_i into disjoint bins $\{B_j \mid j = 1, \ldots, m\}$ such that $\bigcup_j B_j = \Omega_{X_i}$. Every bin B_j is associated with a real number $g(B_j)$ denoting the value of the interval. In real-world applications, the state space of the random variable is unknown but is based on the sample data. The value associated with each bin B_j corresponds to the sample mean of the samples that are included in the bins, $\frac{1}{|B_j|} \sum_{x_i \in B_j} x_i$, in which $|B_j|$ denotes the number of $x_i \in B_j$.

The equal width discretization method partitions the state spaces Ω_{X_i} into bins of equal width. The equal frequency discretization approach divides the samples into quantiles. Both are unsupervised methods and require a parameter specifying the number of bins into which the original state space should be partitioned.

In addition to these two unsupervised discretization methods, four supervised discretization methods are used. First, the entropy error-based approach, dynamic discretization [168] is considered,[4] specifically developed for Bayesian network inference. Second, the minimum description length principle discretization [71] is employed, which iterates through potential cut-points recursively to minimize information entropy with respect to a chosen target variable. Third, ChiMerge [124] is applied, a discretization technique that continuously merges fine intervals based on the χ^2 statistic. Fourth, class-attribute interdependence maximization [133] is used, which discretizes the continuous variables intending to maximize interdependency with the target variable [49]. The latter three supervised discretization methods have been chosen because they performed well on a variety of discretization tasks [77].

Discretization of a continuous Bayesian network is followed by parameter learning, which involves the estimation of the CPTs. In this section, the maximum likelihood

[4]The implementation available at `https://github.com/PCiunkiewicz/dynamic-discretization` is used, adopting the parameter settings deemed most optimal by the implementer.

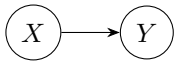
estimate and the Bayesian method with adjusted empirical Bayes type 2 maximum likelihood priors [121, 85] are considered. In the latter, the prior is initially estimated through MLE but refined by substituting 0 probability values with a minimal value (0.0001). This adjusted prior is subsequently used to infer the posterior CPTs with the data. While the maximum likelihood estimates are sufficient to conduct inference on non-causal datasets, the causal datasets require the Bayesian approach to prevent any violations of the positivity assumption as described in Chapter 3 [267]. The differences in results between both methods are discussed together with all the results of the experiments in the Section 5.4.3.

5.2.2 BDD Encoding and Weighted Model Counting

Binary decision diagrams [40] are rooted directed acyclic graphs which represent Boolean functions $f : \{0,1\}^n \to \{0,1\}$, although by storing additional information outside the BDD they can also be used to represent pseudo-Boolean functions $f : \{0,1\}^n \to \mathbb{R}$. Two important properties of BDDs are their ability to compactly represent many functions by identifying redundancies, and their support for efficient operations (i.e. polynomial-time in the size of the BDD), such as computing marginal probabilities.

The joint probability distribution given by a BN is effectively a function of the form $f : \{0,1\}^n \to \mathbb{R}$ and can thus be encoded in a BDD. This is done by encoding each CPT entry in a small Boolean expression, from which a BDD can then be built using primitive BDD operations for logical and ($\wedge$), or ($\vee$), not ($\neg$), etc. As an example, consider the BN given in Figure 5.2. To capture the (integer) values of X and Y, Boolean variables $\{x_0, y_0, y_1\}$ are introduced, while unique probabilities are related to Boolean variables θ_i. As an example of the encoding of specific CPT entry, $P(Y = 2 \mid X = 0) = 0.4$ is encoded as $(\neg x_0 \wedge y_1 \wedge \neg y_0) \Rightarrow \theta_2$, where $\neg x_0$ corresponds to $X = 0$ and $y_1 \wedge \neg y_0$ corresponds to $Y = 2_{\text{dec}} = 10_{\text{bin}}$. The relationship $\mathsf{val}(\theta_2) = 0.4$ is stored outside of the BDD.

Computing marginal or conditional probabilities from a BDD that encodes a joint probability distribution can be done using so-called *weighted model counting* [46]. During weighted model counting the BDD is traversed, relevant probabilities are gathered along the way, and each node is visited at most once, resulting in a computation time linear in the size of the BDD. To compute interventional queries, the *do*-operator has been implemented through the adjustment formula (Equation 5.2) that utilizes the efficiently computed marginal and conditional distributions.

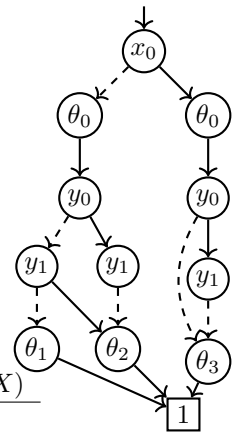


(a) BN structure

$P(X=0)$	$P(X=1)$
0.5	0.5

X	$P(Y=0\|X)$	$P(Y=1\|X)$	$P(Y=2\|X)$
0	0.2	0.4	0.4
1	0.33	0.33	0.33

(b) Conditional probability tables



(c) BDD

θ_i	val(θ_i)
θ_0	0.5
θ_1	0.2
θ_2	0.4
θ_3	0.33

(d) θ_i values

Figure 5.2: An example Bayesian network (a, b) and the corresponding BDD (c). The probabilities associated with the Boolean variables θ_i are shown in (d). In the BDD, solid (dashed) edges correspond to positive (negative) assignments. For clarity, edges to the 0-terminal are omitted in (c).

5.2.3 Optimization

Given the above discretization and encoding mechanisms, interventional queries can be computed and the causal global optimization problem (Equation 5.1) can be addressed heuristically. To illustrate this point, a variety of optimization algorithms from the *black-box optimization* literature [19], implemented in the Nevergrad [190] platform, are selected. The set of algorithms includes: (a) random search as a baseline, (b) two evolutionary algorithms (differential evolution [235] (DE) and (1+1) evolutionary strategy [32] (OnePlusOne)), (c) local search methods (BFGS [39, 73, 84, 215] and Powell's method [186]), (d) the algorithm wizard 'NGOpt39' [160], which automatically selects an optimizer based on problem characteristics such as the number of variables.

5.3 Experimental Setup

In applying the methodology depicted in Figure 5.1 to a range of Bayesian networks, the quality and efficiency of the resulting inference as well as the performance of the optimization algorithms are evaluated. Section 5.3.1 introduces the measures used to evaluate the quality of the discretization or inference and the measure used to assess

the computational cost of inference with decision diagrams. The evaluation procedure for the optimization process is also detailed. The specifics of the causal and non-causal Bayesian networks, including which networks are subjected to which experiments, are outlined in Section 5.3.2.

5.3.1 Evaluation Measures

First, different measures to assess the quality of discretization and inference are discussed, followed by a measure to evaluate the computational costs. Finally, the evaluation of the optimization algorithm is discussed.

Measuring the quality of discretization and inference

While f-divergences measure differences between probability distributions on the same measurable space [210], they are unsuitable for comparing a discretized state space and its continuous counterpart. Instead, the Wasserstein distance is used, specifically the Euclidean first-moment Wasserstein distance or earth mover's distance, to assess discretization quality as it is a common metric to compare (multivariate) distributions [255, 205, 12]. The earth mover's distance quantifies the dissimilarity between two probability distributions by measuring the minimum 'cost' to transform one distribution into the other.

A high-quality discretization does not necessarily imply that a query of interest can be computed accurately. Fortunately, as the synthetic BNs used in the experiments possess specific distributions that permit exact inference methods, access to the conditional inference queries is available. To evaluate inference quality, the conditional expected value of the original Bayesian network ($\mathbb{E}[Y \mid X]$) is compared to its discretized counterpart ($\mathbb{E}_{disc}[Y \mid X]$) using the weighted root mean squared error, where the weights adjust for the probability of the conditioned-on variables. For the causal Bayesian networks, the percentage error of the average treatment effect is used. The reader is referred to Appendix B.3 for a detailed description of these evaluation measures.

Measuring the computational costs of inference

As outlined in Section 5.2, inference using binary decision diagrams reduces to weighted model counting, which takes time linear in the size of the BDD. Therefore, the number of BDD nodes is considered a proxy for the computational costs of inference. Although BDDs can potentially grow exponentially in the size of the Bayesian network, they

typically remain smaller, enabling more scalable inference compared to traditional methods like variable elimination or belief propagation.[5]

The reported inference time for BDDs includes both compilation and weighted model counting. The runtime of inference with traditional methods is compared to the inference time with BDDs. Since VE and BP are implemented in Python and weighted model counting in C++, comparing their runtimes directly is inappropriate. Instead, scalability is assessed by measuring the time speed-up (seconds) as the number of bins in the Bayesian network increases.

Measuring optimization performance

For each optimization algorithm and each Bayesian network, 10 independent runs are performed of 2000 evaluations of the objective function each, where an evaluation consists of calculating the expected value of the outcome variable given an intervention set (see Equation 5.2), which is to be minimized in all networks considered in this paper (see Equation 5.1).

To represent the problem within the optimization algorithms, each node eligible for intervention is assigned a value between 0 and the number of bins if an intervention is performed; for convenience, a negative value is used to indicate the absence of intervention on that node. Within the Nevergrad library, such a representation gets translated to a real-valued one to allow continuous optimization algorithms to tackle this problem as well [190]. To track the optimization performance, IOHexperimenter [63] is used, a benchmarking module from the IOHprofiler environment [246], which allows us to track the optimization process fully. The expected value of the interventional distribution across 2,000 evaluations, averaged over 10 runs is reported.

5.3.2 Bayesian Network Description

In this section, the specifications of the non-causal and causal Bayesian networks that are subject to experimentation are highlighted. As the optimization experiments aim to embed the BDD-inference framework within an optimization loop, the BNs used in these experiments are generally a bit more expansive than those used solely in inference experiments. First, some general statistics for the Bayesian networks are summarized in Table 5.1, followed by a detailed description of each network. Finally, Table 5.2 presents the experimental characteristics of each Bayesian network.

[5]The Python implementation of pgmpy [11] is used for VE and BP to compare runtimes between BDD-inference and traditional inference.

Table 5.1: The general characteristics of all Bayesian networks. These include both synthetic and real datasets, varying in sample sizes, network complexity, and structural properties. The maximum number of parents (in-degree) serves as a proxy for the computational cost of inference. Due to the size of the Mehra and Arth networks and the 64GB RAM constraint, pruned but computationally equivalent versions were used for evaluation [23].

Dataset	Kind	Variants	Samples	Network		Max
				nodes	edges	parents
LG	synthetic	36	100-5000	5	4	2
NM	synthetic	8	100-500	2	1	1
CQ	synthetic	1	2500	3	3	2
Lalonde	real	1	2676	10	17	9
MC	synthetic	1	4000	12	15	6
Arth*	real	1	5000	107	150	17
Toy	synthetic	1	1000	3	2	1
Climate	real	1	293	8	11	3
Mehra*	real	1	5000	24	71	9

Linear Gaussian (LG) Bayesian network. Samples are drawn from a linear Gaussian Bayesian network [175] with random variables X_1, X_2, X_3, X_4, X_5. In total, 36 inference experiments were conducted, varying in sample size (N) and distribution parameters. To ensure a balanced experimental design, Sobol sequences were employed [78]. Detailed experimental specifications are provided in Tables B.3 and B.4 of Appendix B.2. The computational costs in terms of the number of nodes in the BDD is drawn against the WRMSE and against the earth mover's distance in Figure 5.5a, 5.5c and Figure 5.5b, respectively.

Normal mixture (NM) Bayesian network. Samples from a normal mixture Bayesian network are generated using a two-node Gaussian mixture model, for the purpose of conducting inference experiments. In this network, X_1 follows a Bernoulli distribution and $P(X_2|X_1)$ is Gaussian, based on similar experiments by Neil et al. [168]. Details on sample sizes and distribution parameters are listed in Table B.5 of Appendix B.2.

Causal quadratic (CQ) Bayesian network. In the context of the inference experiments, data is sampled from a quadratic data-generating process. The confounder Z is distributed normally and has a quadratic effect on outcome variable Y while also affecting treatment variable T. For the full specifications of the distribution of this experiment [176], the reader is referred to Appendix B.2. The computational costs

of inference have been set out against the percentage error of the ATE in the Pareto front of Figure 5.5d.

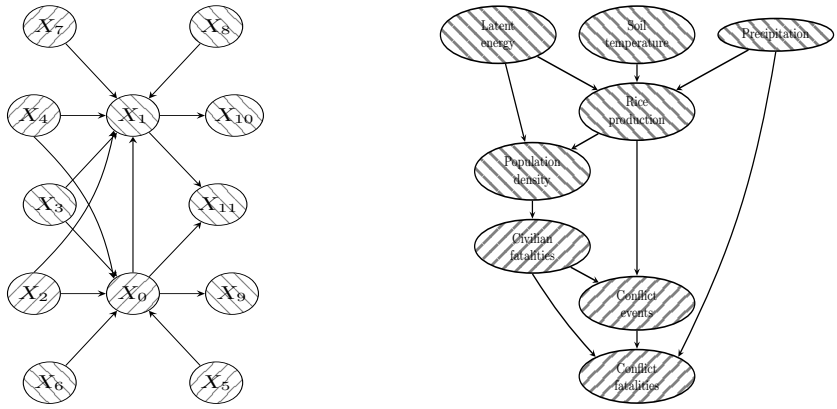
Lalonde causal Bayesian network. The Lalonde causal dataset is a real causal dataset in which the effect of temporary employment on income is studied [135], given confounding variables. Since both an observational [64] and an experimental dataset [135] are available, the non-parametric estimates of the average treatment effect can be compared with the difference in means in the observational and experimental datasets. The comparative analysis of computational costs of inference is presented alongside the percentage error of the ATE in the Pareto front depicted in Figure 5.5e.

Mixed confounding (MC) Bayesian network. To support both discretization and optimization experiments, samples are drawn from a synthetic dataset with mixed confounding, as detailed in the Csuite benchmarking causal datasets [79]. This dataset, depicted in Figure 5.3a, includes both continuous and discrete variables that causally influence multiple nodes in the graph in a non-linear manner. Figure 5.5f presents the computational costs of inference against the earth mover's distance within a Pareto front.

After discretizing the continuous variables into 30 bins, two optimization experiments are conducted on the outcome variables X_{10} and X_{11}, using the minimal intervention set $\{X_2, X_3, X_4, X_5, X_6, X_7, X_8\}$ [141]. Figures 5.6a and 5.6b show that DE and NGOpt19 outperform other optimization algorithms.

Arth Bayesian network. This large Gaussian Bayesian network, sourced from the GeneNet package and featured in bnlearn, contains plant expression data [173]. Due to the enormous size of the networks, a computationally equivalent pruned version of the network is compiled where every node is discretized into 6 bins in order to reduce compilation time [23]. A total of 6 intervention variables have been chosen for the optimization procedure. The results of the optimization experiments are shown in Figure 5.6d.

Toy. This dataset [4] contains a three-node $X \to Z \to Y$ Bayesian network discretized into 100 bins. While not interesting from an optimization perspective, it benchmarks inference quality post-discretization, focusing on possibly-optimal minimal intervention set [141] $\{Z\}$. Most optimization algorithms quickly converge to the



(a) Mixed Confounding BN **(b)** Climate Causal BN

Figure 5.3: Bayesian network structures corresponding to the mixed confounding and climate experiments. The hybrid nodes are distinguished by interior line patterns: dashed lines oriented from the top-left to bottom-right denote nodes with *continuous* values, whereas those from the top-right to bottom-left signify nodes with *discrete* values.

optimal solution in the discretization, $Y = -1.866$, which is near the known exact solution $Y = -1.856$.

Climate. The dataset comprises 294 samples, tapping into the interplay between climate and conflict as depicted in Figure 5.3b. It includes conflict, climate, environmental, and demographic data at the municipal level in southeastern Iraq. An understanding of the variables and the details of the empirically verified causal structure can be found in Malekovic et al. [153].

The outcome variable is the number of conflict fatalities. The interventions in consideration are precipitation, rice production, and population density, although direct intervention may be challenging. Indirect policy measures such as water management and development projects can be targeted to increase water availability and balance demographic distribution, respectively.

Figure 5.6e shows that DE and NGOpt19 are the best-performing algorithms. The tables in Figure 5.6f indicate best-found objective values and intervention values compared to the sample means in the dataset.

Mehra. This conditional linear Gaussian BN from bnlearn [213] explores the correlation between air pollution and health outcomes [247]. Due to its considerable size, the compilation of BDD is restricted to those segments of the network pertinent to

the optimization task [23]. The results can be seen in Figure 5.6c, where the identified best-performing algorithm is random search, closely followed by DE and NGOpt39.

Table 5.2: The experimental characteristics of the Bayesian networks. The table indicates whether each dataset is used for evaluating discretization quality (Disc), inference quality (Inf), or optimization performance (Opt). Inference quality evaluation is only possible when the ground truth of inference queries is available. For these inference experiments, the corresponding evaluation metric is reported. For optimization experiments, the number of discretization bins and the number of intervention variables considered are specified.

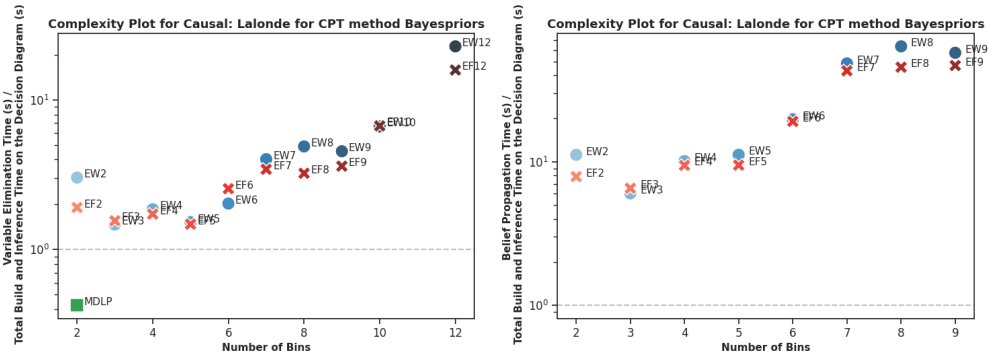
Dataset	Experiments	Inference comparison	Discretization bins	Intervention variables
LG	Disc/Inf	WRMSE	*NA*	*NA*
NM	Disc/Inf	WRMSE	*NA*	*NA*
CQ	Disc/Inf	PE ATE	*NA*	*NA*
Lalonde	Disc/Inf	PE ATE	*NA*	*NA*
MC	Disc/Opt	*NA*	30	7
Arth	Disc/Opt	*NA*	6	5
Toy	Opt	*NA*	100	1
Climate	Opt	*NA*	20	3
Mehra	Opt	*NA*	4	8

5.4 Results

First, scalability results are presented, comparing inference speed using binary decision diagrams against conventional approaches. Subsequently, Pareto fronts show the trade-off between computational cost and the quality of discretization and inference. Finally, the optimization algorithm's performance is analyzed across experiments by examining the average objective function value relative to the number of evaluations.

5.4.1 Scalability of Inference Method

In Figure 5.4, the speedup of Bayesian network inference via binary decision diagrams is compared to inference with variable elimination (VE) (Figure 5.4a) and belief propagation (BP) (Figure 5.4b) for the Lalonde Bayesian network. The Lalonde network has been chosen since it has a relatively high maximum in-degree, a proxy for the computational costs of inference. The fact that inference with binary decision diagrams becomes at least over 10 times faster than VE or BP as the number of bins increases underscores a notable improvement in scalability (in fact, for BP this is true for over 5 bins).



(a) BDD speedup versus VE **(b)** BDD speedup versus BP

Figure 5.4: The speedup plots for using binary decision diagrams as opposed to VE (a) or BP (b) for inference for the Lalonde experiment. The red crosses refer to EF binning, the blue circles represent the EW binning, and the MDLP binning is indicated by a green square. As the number of bins increases, using decision diagrams is more than 10 times as fast as both VE and BP.

5.4.2 Trade-off Computational Cost and Quality of Discretization and Inference

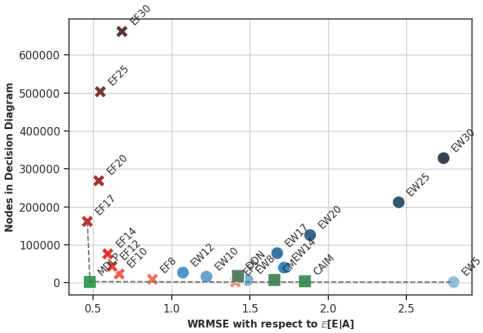
While all Pareto fronts are available at Zenodo,[6] a representative selection across all Bayesian networks and measures is highlighted in Figure 5.5. These Pareto fronts clearly demonstrate that increasing the number of bins results in a reduction of the earth mover's distance but an increase in computational costs. Simultaneously, the WRMSE and the percentage error of the ATE decrease as the number of bins rises, up to a certain number of bins.

To facilitate the interpretation of results across various experiments, the Pareto fronts have been condensed into the heatmaps presented in Tables B.1 and B.2 of Appendix B.1. All the experiments yield the following four key findings.

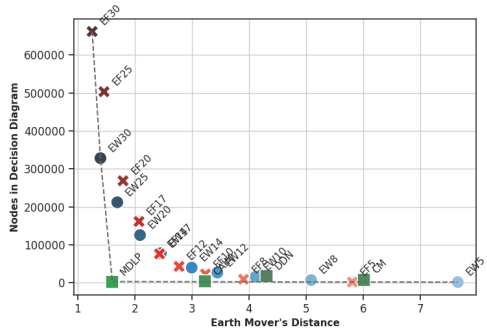
First, the solutions with the lowest earth mover's distance to the original BN are the most-binned solutions as can be observed in Figures 5.5b and 5.5f. In general, it can be observed that the earth mover's distance decreases when the number of bins used to discretize the BN increases, but the distance reduction becomes lower as the number of bins grows larger.

Second, the difference between inference results when estimating the CPTs with maximum likelihood estimate or the Bayesian method with adjusted empirical Bayes type 2 maximum likelihood priors is negligible. This similarity is evident from the

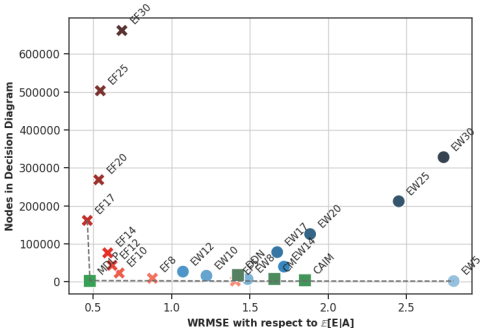
[6] `https://zenodo.org/records/11202314`



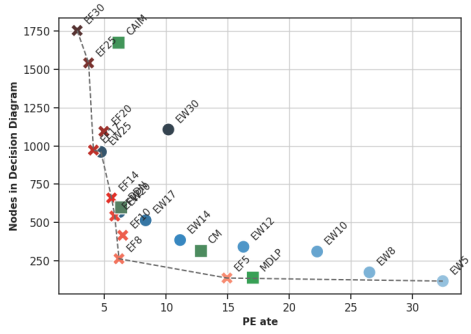
(a) WRMSE for the linear Gaussian experiment 9 with CPT method MLE.



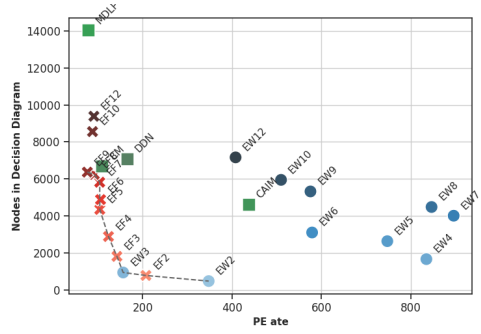
(b) Earth mover's distance for the linear Gaussian experiment 9.



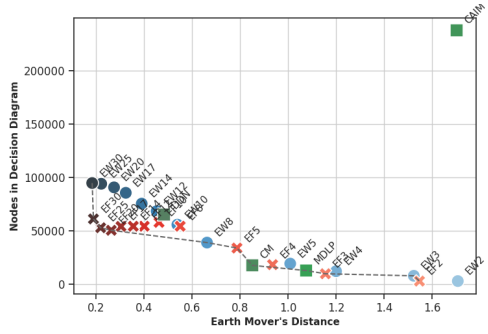
(c) WRMSE for the linear Gaussian experiment 9 with CPT method EBP.



(d) Percentage error of the ATE for the causal quadratic DGP.



(e) Percentage error of the ATE for the Lalonde dataset.



(f) Earth mover's distance for the mixed confounding dataset.

Figure 5.5: Number of nodes in the BDDs versus various evaluation measures for several discretization approaches with different parameter settings, per dataset. Approaches representing trade-offs between objectives (axes, both to be minimized) lie on the Pareto front (dashed line).

plots in Figure 5.5a and 5.5c, and supported by the data in Table B.2 in Appendix B.1. Additional discrepancy plots on Zenodo[7] further illustrate this negligible difference.

Third, the WRMSE and the PE decrease when adding bins up to a certain number of bins whereafter it increases again, indicating overfitting in data-sparse areas of the root variable. The Pareto fronts of Figure 5.5c, 5.5d, and 5.5e show that the bending point differs per experiment. Generally, more available samples or simpler BN structures lead to the solution with the lowest error being often a more intensely-binned solution.

Finally, it can incidentally be observed that one of the supervised discretization methods dominates the other solutions (as in the case with CM and MDLP in Figure 5.5f). However, no supervised discretization method performs exceptionally well across all experiments on the considered measures.

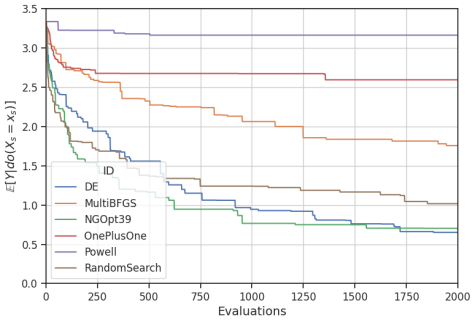
5.4.3 Optimization Performance

This section evaluates the performance of the optimization algorithms. Results from the previous section indicate that high-binned equal frequency discretizations yield among the most accurate query results. Therefore, these discretizations are used, reducing the bin count only when required to stay within the 64GB RAM constraint during BDD compilation. Details about the number of bins in the discretization and the number of interventional variables are displayed in Table 5.2, while the results are presented in Figure 5.6.

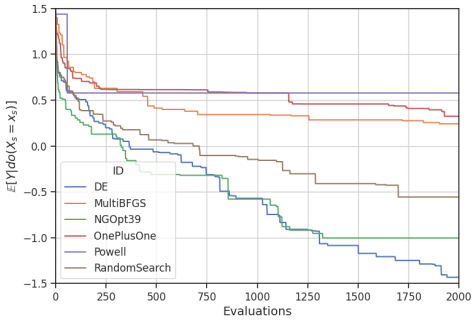
The more local methods (BFGS, Powell, OnePlusOne) perform rather poorly. Only DE and NGOpt are able to outperform the random search baseline, suggesting that the underlying optimization problem might be multimodal. This might also be connected to the choice of internal problem representation selected from Nevergrad, as working directly on the discrete variables might be more suitable for these local search methods. While this points to a need for further examination of the specifics of the optimization procedure, the results nevertheless illustrate that in general these problems *contain sufficient structure* that heuristic black-box optimizers can improve over the performance of random search.

As can be observed from Table 5.6f, with interventions tailored to increasing rice supply/production, equitable population management leading to a more balanced demographic distribution, and increased precipitation or related water management interventions, the expected value of conflict fatalities can be reduced by 85.9% with

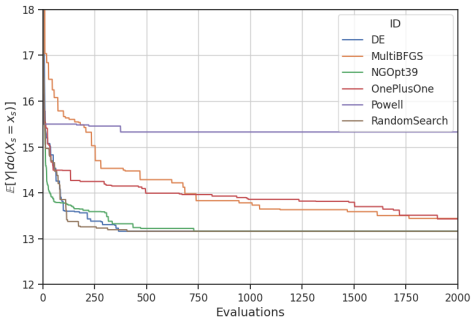
[7] `https://zenodo.org/record/8211601`



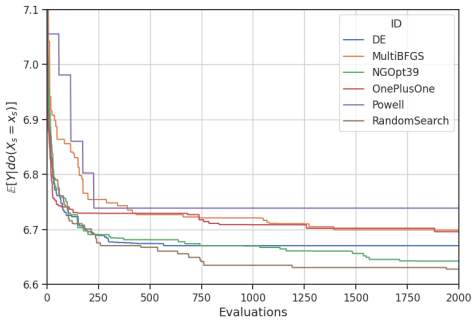
(a) Mixed Confounding: $Y = X_{10}$



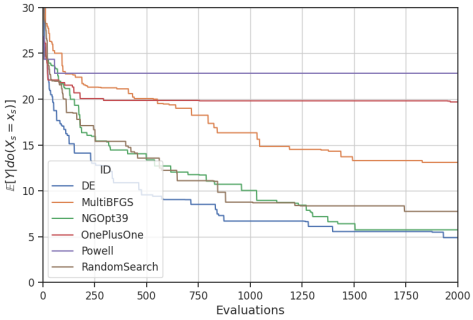
(b) Mixed Confounding: $Y = X_{11}$



(c) Mehra



(d) Arth



(e) Climate

	$\mathbf{y}^*$	$\bar{\mathbf{y}}$
Conflict Fatalities	4.89	34.6

	$\mathbf{x}^*$	$\bar{\mathbf{x}}$
Rice Production	36.9	13.0
Population Density	215	811
Precipitation	25.0	19.8

(f) Climate Dataset: Best-found solutions

Figure 5.6: The best-found expected value of the interventional distributions of 6 optimization algorithms over 2000 evaluations, averaged over 10 runs for the 4 considered datasets (a-e). For the climate dataset, the table in (f) corresponds to found objective $\mathbf{y}^*$ and intervention values $\mathbf{x}^*$ compared to sample mean values $\bar{\mathbf{y}}$, $\bar{\mathbf{x}}$.

respect to the mean.

5.5 Conclusion and Future Work

This chapter proposes a methodology to optimize causal interventions in hybrid causal Bayesian networks using only observational data. The methodology consists of discretizing the hybrid Bayesian network and encoding it into a binary decision diagram. Once the binary decision diagram is compiled, query costs become negligible, allowing the deployment of heuristic optimization algorithms that optimize for the optimal intervention. Given that discretization of a Bayesian network entails information loss, benchmarking this approach against established approximate inference methods, such as sampling or variational techniques, constitutes a promising direction for future research [127]

The estimates in Table 5.6f demonstrate the practical potential of the methodology, though they rely on a preliminary causal structure under the i.i.d. assumption introduced in Chapter 3. Chapter 6 further shows that causal estimates remain attainable when this assumption is relaxed. Extending the proposed methodology to account for causal spillovers or multiple strategic actors would enhance its relevance in complex, real-world scenarios.

Chapter 6

Real-World Applications

The preceding chapters introduced key concepts from causality and game theory, alongside a methodology for optimizing causal interventions in hybrid Bayesian networks.

This chapter applies these theoretical foundations to the context of complex security environments through two specific case studies. First, causal techniques are applied to uncover a causal structure and estimate causal effects related to environmental conflict in Iraq in Section 6.1. Second, causal game-theoretic concepts are utilized to simulate the effectiveness of countermeasures against hybrid threats in Section 6.2. Environmental conflict and hybrid threats are examined as illustrative examples of complex security environments, as they involve interdependent challenges, diverse threat actors, and evolving risks that contribute to unpredictable and dynamic security conditions.

In doing so, this chapter addresses RQ3: *How can the proposed (strategic) causal concepts be applied to complex security environments?* The content closely reflects two peer-reviewed journal articles [250, 154], to which the reader is directed for further detail.

6.1 Environmental Conflict

Despite significant advances in conflict research methods [188, 238, 132, 44], the outbreak of armed conflict remains difficult to predict, largely due to unresolved questions about its underlying causes. While environmental security research has explored specific causal pathways linking environmental factors to conflict [66, 103, 106, 20, 199,

209], most studies rely on observational data and confirm only limited mechanisms. However, the randomized controlled trials discussed in Chapter 3 are neither practical nor ethical in the context of armed conflict. Therefore, more comprehensive causal explanations remain elusive, leaving a critical methodological gap that hinders both scholarly understanding and effective policy interventions.

In addressing this gap, this section applies the concepts introduced in Chapter 3 to infer causality from non-experimental observations of armed conflict. Using causal discovery and inference methods, commonly hypothesized causal pathways are tested. The considered cross-section consists of 294 non-experimental observations, one for each subdistrict in Iraq (Arabic: nawāḥī) as the unit of analysis. The outcome variable is the count of conflict events. Each observation is additionally described in terms of explanatory variables, including demographics, vital resources, environment, and weather. These observations were sampled from several geocoded maps [188, 208, 164, 159, 50, 114, 1]. From these observations, an empirical causal mechanism of linkages between environment and conflict was retrieved. Represented as a causal graph, the mechanism shows causal pathways from environmental variables to armed conflict outcomes. The mechanism is characterized by causal effects of these variables on the count of conflict events, accounting for causal spillover wherever these effects could be identified and estimated.

Sections 6.1.1 and 6.1.2 present the hypotheses and detail the data and methodological approach. Section 6.1.3 reports the empirical findings, while Section 6.1.4 concludes with a discussion of the implications and potential directions for future research.

6.1.1 Hypotheses

This section derives the causal hypotheses discussed below from relevant findings in the literature. Adapted from Sakaguchi, Varughese, and Auld [209], Figure 6.1 outlines a hypothesized mechanism through which environmental factors may contribute to conflict, as extensively discussed in environmental security literature. The hypothetical linkages are rooted in environmental causes. The figure respectively distinguishes between direct and indirect linkages (i.e., paths A and B, respectively). The latter are indirect because environmental scarcity hypothetically plays a mediating role between environmental causes and armed conflict outcomes.

Longer-term weather patterns have been argued to cause armed conflict directly [112, 111, 225, 81, 21]. The direct link between long-term weather patterns and armed

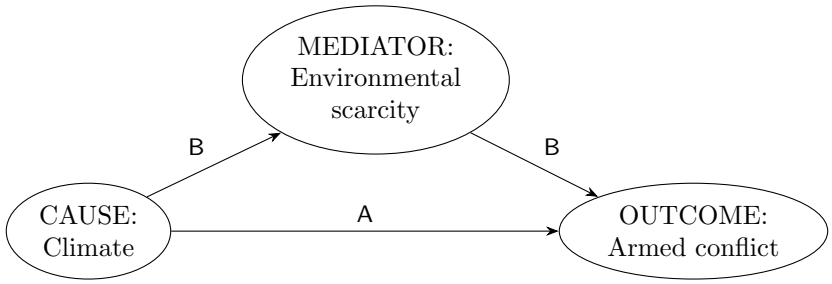


Figure 6.1: Hypothesized causal pathways originating from linkages between environment and conflict.

conflict can be seen in how populations respond to environmental changes. For instance, environmental disruptions affecting livelihoods can prompt community mobilization or lead armed groups to intervene, anticipating adverse outcomes. These actions may lead to direct causal effects of environmental factors to armed conflict. Such effects can originate from soil moisture, temperature, or simply the fact that different physical surroundings absorb or release accumulated heat differently (i.e., also referred to as latent heat or energy) [111, 209]. For example, droughts in East Africa have increased cattle raiding and inter-communal violence among pastoralist groups [72]. Therefore, it is hypothesized that changes in soil moisture, temperature, and latent energy directly cause changes in armed conflict activity (H1).

- H1 a): An increase in latent energy in the form of heat directly causes an increase in armed conflict activity.
- H1 b): An increase in skin temperature directly causes an increase in armed conflict activity.
- H1 c): An increase in soil moisture directly causes a decrease in armed conflict activity.

Further, environmental processes have also been argued to cause armed conflict indirectly [16, 209, 131]. Causal mediation of environmental effects on armed conflict concerns primarily scarcity of vital resources, also referred to as environmental scarcity [103, 104, 105, 129]. Environmental scarcity has been argued to mediate the environmental effects on armed conflict [209, 105]. For instance, land scarcity in Chiapas, Mexico has been hypothesized to mediate environmental pressures into insurgency and civil violence [105]. While causal mediation is elaborated in more detail below,

it can already be hypothesized that the effects of environmental processes indirectly cause armed conflict.

- H2 a): An increase in latent energy indirectly causes an increase in armed conflict activity.
- H2 b): An increase in skin temperature indirectly causes an increase in armed conflict activity.
- H2 c): An increase in soil moisture indirectly causes a decrease in armed conflict activity.

To refine the understanding of indirect causal mechanisms, causal mediation analysis can incorporate specific conditions through which environmental effects manifest. Agricultural and pastoral systems have been shown to shape societal responses to long-term climatic variability, including migration [259]. Land degradation, desertification, and water scarcity are identified as mediators of environmental impacts on violent conflict [98, 72, 116]. Wheat production, a key agricultural output in Iraq [45], has been found to mediate temperature effects on violence [145].

Demographic factors, such as population size, growth, density, and migration, have also been linked to conflict [241, 187, 192]. Resource scarcity is argued to affect denser populations more acutely, increasing the likelihood of conflict under environmental stress [16, 2]. These mediating pathways are captured in the following hypotheses.

- H3 a): Given the indirect paths from the environmental processes to armed conflict activity, wheat production causally mediates the indirect effects of envi-

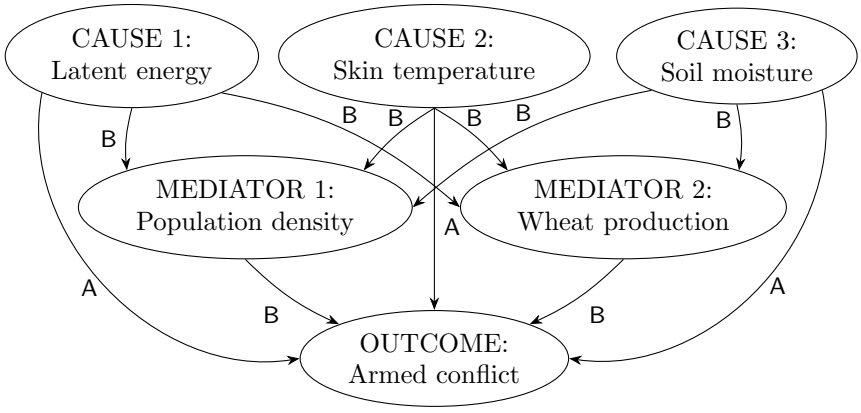


Figure 6.2: The hypothetical causal structure of all hypotheses combined.

ronmental processes on armed conflict activity, by causing an additional decrease in armed conflict activity.

- H3 b): Given the indirect paths from the environmental processes to armed conflict activity, population density causally mediates the indirect effects of environmental processes on armed conflict activity, by causing an additional increase in armed conflict activity.

With these hypotheses, it is possible to compose the entire hypothetical causal structure of linkages between environment and conflict, as shown in Figure 6.2. The figure again distinguishes between direct and indirect linkages, corresponding to edges A and B, respectively. As can be seen in the figure, grounded in environmental processes, the scarcity of vital resources exposes the population to existential stress. Both the density of population and the scarcity of agricultural resources aggravate these effects.

6.1.2 Data and Methods

Data

The units of analysis are all Iraq's 294 subdistricts from January 1, 2020, to January 1, 2022, aggregated as a single cross-section. Data included conflict events, such as violence against civilians and battles from ACLED [188], where total conflict events—aggregated across battles, explosions, violence against civilians, protests, riots, and strategic developments—served as the outcome variable. Environmental and demographic explanatory variables were sourced as geo-coded grids from the Humanitarian Data Exchange, ECMWF's ERA5-Land [164], NASA [114, 208, 159], CIESIN at Columbia University [50], and MapSPAM [232]. Skin temperature, representing the interface temperature between the earth's surface and atmosphere, was selected due to its relevance for agriculture and water availability [122]. Soil moisture at 28–100 cm depth, which impacts crop viability and water access, was also included. Latent energy, measuring heat flux linked to evaporation and condensation, was used to capture broader hydrometeorological dynamics [208]. As a proxy for environmental scarcity, wheat production data were retrieved from MapSPAM, given wheat's importance to Iraqi agriculture [232, 145]. Population density, previously linked to conflict vulnerability, was sourced from CIESIN's Gridded Population of the World dataset [50]. Due to data scarcity in Iraq, other societal and political variables commonly used in conflict research were unavailable at an appropriate resolution [261, 3].

Methods

By applying causal methodology to non-experimental observations, causal paths and effects behind the causal mechanism of such linkages can be disentangled and quantified. As the methodological concepts are introduced in Chapter 3, the discussion here is limited to the specification and implementation of the introduced methods.

Since there are a limited number of observations, GES was selected as the causal discovery algorithm as it has been found appropriate in simulation studies involving small sample sizes [155]. The considered loss function was the Bayesian information criterion. The output of GES was the likeliest DAG, given the observations. The nodes of the DAG correspond to the armed conflict activity and explanatory variables. The edges correspond to respective causal relationships between them [155]. Causal estimands of the explanatory variables were identified using the backdoor criterion and adjustment formula. Acknowledging that the observations in question are spatially correlated in the sense that the climatological variables of one municipality may have an influence on the climate-conflict dynamics of another municipality, the estimation procedure invokes SESEM to account for spatial confounding [137].

A straightforward way to make this spatial confoundedness explicit is to use distances between municipalities. Distances are modeled by computing municipal centroids and the shortest paths connecting them. The first step of applying SESEM is fitting an initial non-spatial SEM to the data [136], assuming independence of errors. Then, spatially explicit variance–covariance matrices are computed for a series of lag distances divided into bins corresponding to distance ranges. Each bin contains 500 data pairs to ensure sufficient sample size for meaningful inference. Inference focuses on the lowest 20% of distances, as spillover effects are more likely in nearby municipalities. Finally, SEM models are fitted for each chosen lag distance, and edge coefficients, standard errors, and p-values are computed. In addition, parameters are defined for individual causal paths such that path-specific coefficients, standard errors, and p-values can be computed.

6.1.3 Results

Empirical causal structure

Section 6.1.1 outlined the hypothetical causal structure linking environmental factors to conflict. Figure 6.3 presents the empirically derived causal structure based on the available non-experimental observations.

Albeit somewhat less expressive, the empirical causal structure largely corresponds with the hypothetical one in Figure 6.2. The conflict nodes cluster together. The only node with only incoming edges is total conflict events. Further, the structure is rooted in environmental processes. Apart from the direct causal path from the temperature node to battle events, all the other paths from the environmental processes to conflict events are indirect.

Because the population density, temperature and wheat production nodes have the highest number of incoming and outgoing edges, these nodes are pivotal to the connectedness of the empirical causal structure. This lends credence to the causal mediation of environmental processes on the conflict outcomes. In fact, without population density and wheat production, the environmental causes would be largely disconnected from the outcomes. Rather than treating this evidence as conclusive, the empirical structure is further used to conduct hypothesis testing.

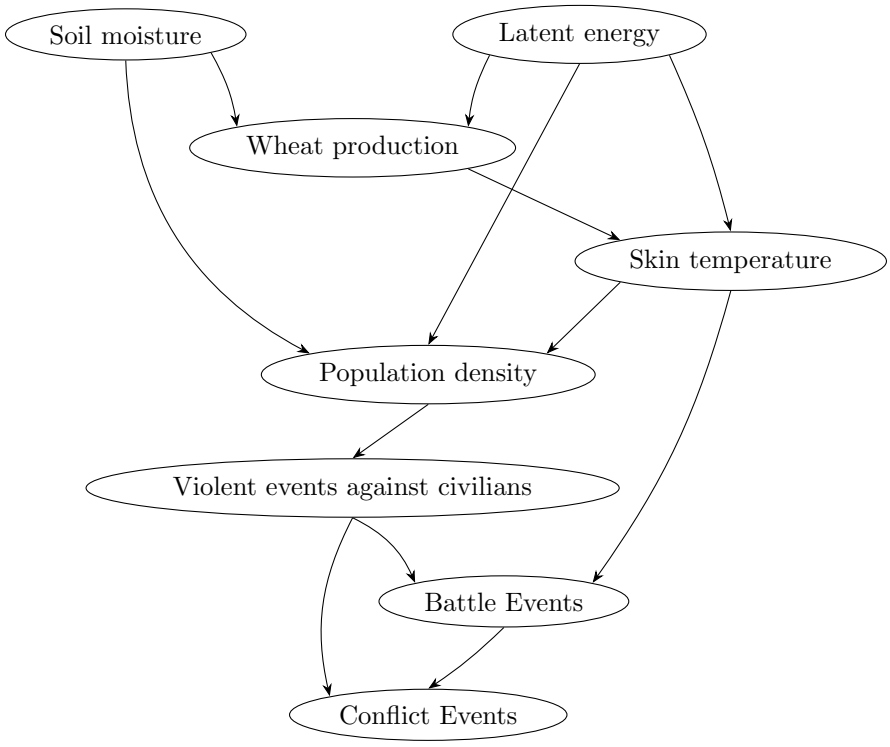


Figure 6.3: The empirical causal structure retrieved from the GES algorithm.

Causal hypotheses

The empirical causal structure can assist with the validation of the causal hypotheses of naturally caused armed conflict. With this structure, a spatially explicit structural equation model was fit. Continuous explanatory variables were z-score standardized, while count variables were left unstandardized to maintain policy-relevance. Causal effects of the explanatory variables on the count of conflict events were estimated.

The SESEM model showcased acceptable performance, with the comparative fit index (ranging 0-1) exceeding 0.9 and the standardized root mean square residual falling below 0.08 for almost all spatial distances [211]. These values are indicators of an acceptable model fit, underscoring the model's effectiveness in capturing the underlying data structure. While the root mean square error of approximation values are occasionally above 0.10, suggesting room for improvement, this is likely attributable to the limited number of observations available.

For each causal estimate, a hypothesis test was conducted to discern whether to attribute the estimate to a random error or not. Table 6.1 lists the causal estimates, their standard errors, and statistical significances.

The first set of hypotheses states that effects of environmental processes directly cause armed conflict outcomes. The changes in latent energy (H1 a), skin temperature (H1 b), and soil moisture (H1 c) were hypothesized to cause a change in the count of conflict events directly. The only environmental process with a direct causal path to conflict events is skin temperature. This path is mediated via battle events and the estimates, standard errors, and statistical significances can be estimated for each of the isolated paths. The results are in Table 6.2. Given the causal structure, the estimated causal effect of skin temperature on total conflict events mediated by battle events is positive for all spatial distances and 47.35 for the non-spillover fitted model, with at least a 0.1% statistical significance level. Since there is no evidence that warrants rejecting the null hypotheses relative to H1 a) and H1 c), hypothesis H1 b) is accepted and hypotheses H1 a) and H1 c) are not accepted.

Further, the second set of hypotheses states that causal effects of the environmental processes on conflict events are mediated. Latent heat was hypothesized to increase the count of total conflicts indirectly (H2 a). Since all effects from latent heat to conflict events are mediated, and the estimated effect of latent heat on the conflict events is positive and 0.1% statistically significant for all spatial distances, hypothesis H2 a) is accepted. Furthermore, soil moisture (H2 c) was hypothesized to cause a decrease in the count of conflict events indirectly. As there are no direct paths from

Table 6.1: Standardized causal estimates across distance bins with standard errors in brackets. All results are statistically significant at the 0.1% level.

Distance	Wheat Production	Latent Heat	Soil Moisture	Skin Temperature	Population Density
0 km	14.02 (3.9)	24.27 (6.2)	-18.30 (4.6)	51.67 (9.5)	29.66 (5.8)
3–27 km	23.48 (4.1)	33.92 (5.4)	-20.72 (4.2)	64.16 (7.6)	22.61 (4.3)
27–40 km	10.69 (2.6)	21.83 (4.3)	-17.93 (3.1)	48.17 (7.4)	29.42 (4.5)
40–50 km	15.62 (3.4)	26.64 (5.2)	-19.95 (3.8)	56.04 (8.1)	26.65 (4.5)
50–59 km	12.38 (2.7)	23.22 (4.4)	-18.93 (3.4)	45.23 (6.5)	34.42 (4.7)
59–66 km	7.59 (2.2)	15.05 (3.5)	-10.41 (2.4)	36.99 (7.0)	16.52 (3.3)
66–73 km	13.99 (3.1)	28.43 (5.4)	-16.99 (3.8)	58.17 (7.8)	18.17 (4.1)
73–80 km	14.62 (3.7)	28.87 (5.5)	-19.53 (4.0)	71.40 (9.0)	32.45 (5.9)
80–86 km	11.18 (2.6)	15.61 (3.6)	-14.71 (2.9)	41.21 (7.0)	26.29 (4.0)
86–92 km	14.45 (2.7)	20.63 (4.1)	-16.43 (3.1)	47.49 (6.3)	23.16 (3.8)
92–99 km	12.00 (2.5)	21.88 (4.3)	-19.68 (3.5)	41.15 (6.0)	26.86 (4.0)
99–104 km	18.30 (3.7)	26.26 (5.2)	-25.15 (4.3)	63.40 (8.3)	43.84 (5.6)
104–110 km	23.17 (4.6)	37.04 (6.6)	-24.85 (4.8)	81.29 (9.4)	27.69 (5.1)
110–116 km	7.70 (2.0)	13.16 (3.1)	-14.46 (2.6)	30.19 (5.9)	30.58 (3.7)
116–121 km	11.67 (2.7)	23.80 (4.8)	-14.44 (3.0)	45.79 (7.4)	29.06 (4.3)
121–126 km	6.92 (1.7)	13.87 (3.0)	-16.06 (2.7)	33.26 (5.4)	27.72 (3.4)
126–132 km	16.08 (3.0)	23.86 (4.5)	-16.46 (3.2)	50.07 (6.8)	21.62 (3.5)
132–137 km	8.66 (2.3)	15.39 (4.0)	-13.12 (2.8)	33.65 (6.8)	18.19 (3.4)
137–142 km	11.02 (2.5)	16.71 (3.8)	-14.94 (2.9)	41.28 (7.0)	22.61 (3.6)
142–147 km	15.89 (3.2)	20.92 (4.6)	-17.00 (3.6)	54.80 (7.8)	31.78 (4.6)
147–152 km	9.75 (2.2)	18.46 (4.0)	-11.33 (2.5)	30.66 (5.7)	15.23 (3.5)
152–158 km	11.11 (2.5)	17.22 (4.0)	-14.67 (3.0)	34.63 (6.3)	20.02 (3.3)
158–163 km	14.61 (3.0)	27.24 (5.2)	-13.79 (3.3)	56.28 (7.8)	25.70 (4.7)
163–168 km	14.06 (2.8)	22.30 (4.6)	-18.70 (3.6)	43.42 (6.9)	29.26 (4.0)
168–173 km	11.76 (2.7)	18.74 (4.5)	-15.62 (3.3)	32.77 (6.6)	28.39 (4.2)
173–178 km	13.12 (3.0)	23.36 (4.8)	-15.54 (3.3)	54.18 (7.6)	24.45 (4.3)
178–183 km	8.16 (2.4)	12.68 (3.4)	-12.00 (2.6)	41.87 (8.0)	24.57 (4.2)
183–188 km	17.17 (3.4)	22.95 (5.0)	-20.85 (4.1)	48.57 (7.6)	39.79 (5.4)
188–193 km	14.64 (2.9)	25.01 (5.1)	-20.75 (4.1)	50.37 (6.9)	19.87 (3.7)
193–198 km	15.00 (3.2)	23.31 (4.9)	-17.27 (3.6)	48.33 (7.7)	24.65 (4.8)

soil moisture to conflict events, their causal effects on conflict events can only be indirect. Given the indirect paths from soil moisture to armed conflict activity, ceteris paribus, a one unit increase in soil moisture causes a decrease in the counted conflict events at all spatial bins, including a -18.30 decrease for the non-spillover fitted model. Therefore, hypothesis (H2 c) is also accepted. Isolating the indirect paths from the skin temperature (H2 b) to armed conflict activity via population density, ceteris paribus,

Table 6.2: Causal effects of skin temperature on conflict across distance bins. Values are standardized estimates with standard errors in brackets. Asterisks denote statistical significance at 5% (*), 1% (**), and 0.1% (***) levels.

Distance	Temperature Population Civilians Battles Conflicts	Temperature Battles Conflicts	Temperature Population Civilians Conflicts
0 km	1.39 (0.83)	47.35*** (9.36)	2.93* (1.25)
3–27 km	0.13 (0.17)	63.34*** (7.50)	0.69 (0.72)
27–40 km	1.29* (0.65)	44.88** (7.31)	2.00** (0.82)
40–50 km	0.68 (0.52)	52.46*** (8.03)	2.91** (1.02)
50–59 km	0.52 (0.42)	41.22*** (6.28)	3.49** (1.28)
59–66 km	0.19 (0.26)	35.14*** (6.93)	1.66* (0.70)
66–73 km	0.87 (0.54)	55.63*** (7.76)	1.66*** (0.54)
73–80 km	0.58 (0.52)	68.62*** (8.88)	2.20* (1.04)
80–86 km	0.30 (0.48)	37.21*** (6.74)	3.70*** (1.18)
86–92 km	0.36 (0.27)	45.67*** (6.25)	1.46 (0.75)
92–99 km	2.05** (0.74)	36.79*** (5.86)	2.30*** (0.64)
99–104 km	1.31 (0.71)	58.73*** (8.06)	3.36* (1.41)
104–110 km	0.58 (0.56)	77.91*** (9.29)	2.79** (1.06)
110–116 km	2.15** (0.76)	24.85*** (5.65)	3.19*** (0.97)
116–121 km	0.95 (0.52)	42.72*** (7.28)	2.12* (0.98)
121–126 km	1.73* (0.70)	28.14*** (5.34)	3.39*** (1.01)
126–132 km	1.13* (0.51)	42.72*** (7.28)	1.82** (0.69)
132–137 km	0.82* (0.39)	30.58*** (6.75)	2.25*** (0.76)
137–142 km	1.53* (0.61)	37.31*** (6.87)	2.45*** (0.74)
142–147 km	1.58 (0.72)	49.34*** (7.62)	3.88*** (1.10)
147–152 km	0.35 (0.24)	29.54*** (5.69)	0.76 (0.49)
152–158 km	0.97* (0.45)	31.38*** (6.18)	2.28*** (0.71)
158–163 km	0.63 (0.51)	53.10*** (7.72)	2.55*** (0.90)
163–168 km	1.04 (0.56)	39.63*** (6.68)	3.04** (1.14)
168–173 km	0.71 (0.48)	30.67*** (6.47)	1.39 (0.83)
173–178 km	1.96* (0.87)	49.37*** (7.56)	2.85*** (0.78)
178–183 km	1.14 (0.65)	38.80*** (7.92)	1.93* (0.82)
183–188 km	1.84* (0.84)	43.89*** (7.36)	2.84** (1.08)
188–193 km	0.52 (0.50)	47.06*** (6.84)	2.79*** (0.85)
193–198 km	1.79* (0.90)	44.92*** (7.56)	2.12** (0.74)

computing causal effects did not yield significant results for all the spatial distances as can be observed in Table 6.2. Therefore, hypothesis H2 b) is not accepted.

Finally, the third set of hypotheses states that causal effects of environmental conditions on conflict events are agriculturally and demographically mediated. Whereas wheat production was hypothesized to cause a decrease in the count of conflict events

(H3 a), population density was hypothesized to cause an increase (H3 b). Given the indirect paths from soil moisture and latent energy to armed conflict activity, ceteris paribus, a one unit increase in wheat production causes an increase in counted conflict events for all spatial distances, including a 14.02 unit increase at 0.1% statistical significance level for the non-spillover model. Given the indirect paths from soil moisture and latent energy to armed conflict activity, ceteris paribus, a one unit increase in population density causes a 29.66 increase in the number of counted conflict events at 0.1% statistical significance level for the non-spatial model and similar significant positive estimates for other spatial distances (see Table 6.1). This evidence leads to the rejection of the null hypotheses relative to the H3 a) and H3 b) hypotheses, whereas hypothesis H3 a) is rejected and H3 b) is accepted at all spatial bins.

6.1.4 Conclusion and Future Work

Armed conflict research is advanced by applying causal methodology to non-experimental data, enabling the retrieval of empirical causal structures, identification of causal paths, and estimation of effects. The research confirms that environmental processes, particularly soil moisture and latent energy, affect armed conflict through demographic and agricultural mediators, addressing key gaps in environmental security literature. Finally, the findings support the design of targeted policy interventions by identifying causal mechanisms and mediators amenable to strategic action. In the context of environmental security, this enables preventive strategies that interrupt causal paths from environmental stressors to conflict outcomes, such as reducing population density through social or migration policies, or mitigating environmental pressures via investments in hydrological infrastructure.

Future research should assess the robustness of the causal findings by relaxing the causal sufficiency assumption and testing for sensitivity to unobserved confounders, particularly social and political variables such as power-sharing, intergroup animosities, and horizontal inequality [261, 3, 188]. This can be achieved using existing causal frameworks designed to account for unobserved confounding [33, 35, 140].

Additionally, the current analysis assumes linear structural relations in the SEM, justified by the need for interpretability, estimation stability, and suitability as a first-order approximation in data-constrained settings. Future research should extend this by applying more flexible, non-linear SEMs to test the robustness of findings and uncover potentially richer causal dynamics.

6.2 Hybrid Threats

Hybrid threats, defined as the coordinated use of violent and non-violent means to exploit vulnerabilities and influence adversaries below the threshold of armed conflict, pose an escalating challenge in an era of growing global interconnectedness. In response, states have implemented a broad spectrum of potential counter-hybrid measures, including economic sanctions, cyber defense strategies, information campaigns, and diplomatic initiatives. However, the effectiveness of these measures remains uncertain due to the complex and opaque nature of hybrid threats, which often operate across multiple domains and take place below the threshold of detection and attribution [125]. Therefore, researchers have resorted to modeling approaches.

While attempting to model hybrid threat dynamics, some authors have turned to game theory to examine strategic interactions among rival states in an effort to overcome the paucity of information available [25, 18]. Others have incorporated scarce data sources into Bayesian modeling techniques [60, 24] with the aim of refining domain knowledge with available data. Although current game-theoretic approaches struggle to capture the complexities and uncertainties inherent in hybrid threat dynamics, Bayesian modeling techniques, while effective at handling uncertainty, fall short in representing strategic interactions.

This section proposes an integrated probabilistic and game-theoretic framework to assess counter-hybrid threat measures under conditions of deep uncertainty. Drawing on influence diagrams to model uncertainties in threat detection, attribution, and mitigation as probabilistic relations [108], the approach is extended to a multi-agent setting [128], enabling the inclusion of strategic interactions. The model captures both cognitive and psychological aspects of deterrence through probabilistic reasoning [30], and strategic decision-making via game-theoretic structures. The interaction between two state-like agents is modeled such that the defender's pay-off reflects the trade-off between the costs of countermeasures and the potential damage from hybrid attacks, whether successfully deterred or not. Optimal countermeasures are identified by maximizing expected pay-offs, while strategic equilibria are derived from the adversary's strategic responses.

In order to test the modeling approach, a cyber threat scenario on critical infrastructure was developed, inspired by real-world incidents. Policy experts and available literature were consulted to identify relevant countermeasures to this cyber threat and collate estimates of the cost, damage mitigation ability, and deterrence ability of each of the counter-hybrid measures. These estimates provided a basis for analyzing counter-

hybrid measures and allowed for gauging their effectiveness across different scenarios, including scenarios where the adversary engages in strategic competition. To validate the proposed approach, the findings were contextualized within the framework of existing studies, and sensitivity analyses were conducted to identify and quantify the most influential variables driving the model's outcomes.

While Section 6.2.1 outlines the methodology for the models considered and elicitation of expert knowledge, Section 6.2.2 highlights the experimental design of the considered hybrid threat scenario. The results are introduced in Section 6.2.3, and some concluding remarks are given in Section 6.2.4.

6.2.1 Methodology

This section outlines the underlying mechanism of the simulation model, along with the process of eliciting inputs required to run simulations using the model. The full scope of the proposed method, including the extent of expert involvement, is illustrated in Figure 6.4.

The model considers the behavior of two agents possessing the characteristics of sovereign states, following the two-agent approach of Balcaen et al. [25] and Attiah et al. [18]. On one side, agent A aims to pursue its strategic objectives using hybrid attacks. On the other side, agent B wishes to protect its national interests and deter

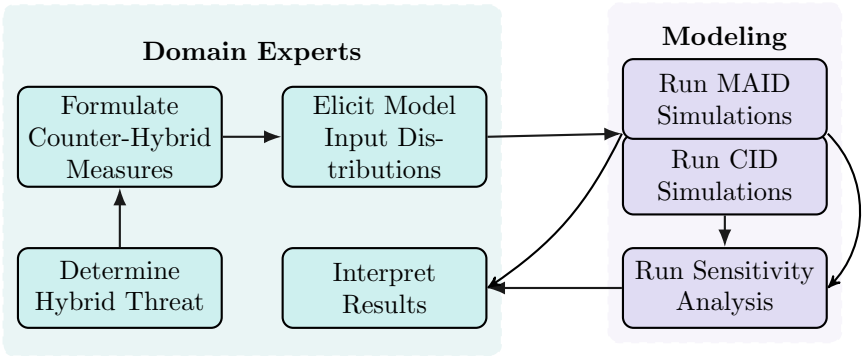


Figure 6.4: The figure illustrates the processes involved in counter-hybrid threat analysis. Initially, domain experts identify the hybrid threat and develop corresponding counter-hybrid measures. They also provide key input parameters, which are used to construct probabilistic input distributions for the model. Samples from these distributions are used to run simulations with the causal influence diagram (CID) as well as the multi-agent influence diagram (MAID) model. Finally, a sensitivity analysis is performed and the model results are interpreted and compared with existing studies.

and defend against hybrid attacks. Agent B, the defender, chooses a counter-hybrid posture to deter or dissuade agent A from carrying out a hybrid operation. For this reason, the strategy is referred to as a counter-hybrid measure. To this purpose, agent B explores available counter-hybrid measures to dissuade the adversary from carrying out hybrid attacks by altering the cost-benefit calculus [38]. The defender may also adopt measures - such as the enhancement of detection and/or attribution capabilities [212] - that would boost resilience and mitigate the potential impact of hybrid conducts [82].

Both the counter-hybrid measure and the hybrid attack bear direct costs. Such costs represent not only the resource costs but also costs involving, for instance, political capital to rally domestic and international support as well as potential costs associated with escalation. The interaction between agents A and B is of a zero-sum nature. Probabilities are used to reflect the considerable degree of uncertainty over the value of key variables that lead to different outcomes. Examples are uncertainty associated with the impact of counter-hybrid measures on the strategic calculus of agent A, as well as with detection, attribution, and the mitigatory impact of the counter-hybrid measure.

First, a causal influence diagram approach is introduced, enabling the optimization of counter-hybrid deterrence strategies when the adversary's responsiveness is estimated probabilistically. This approach is then extended into a multi-agent influ-

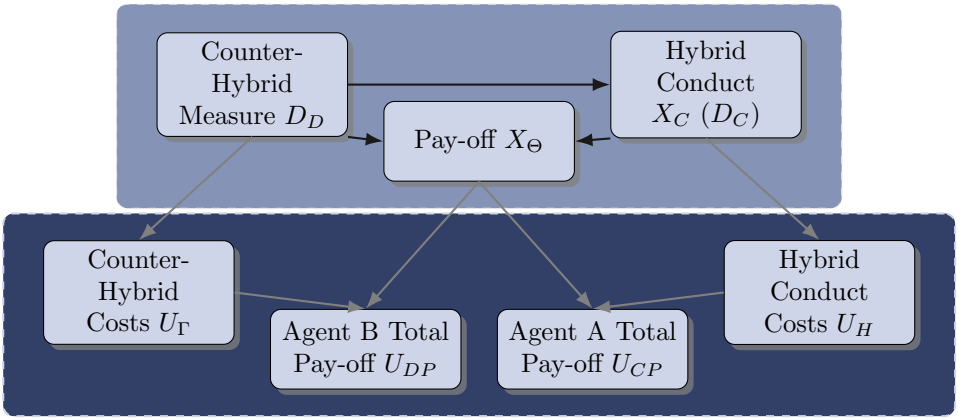


Figure 6.5: (Multi-agent) influence diagram encoding hybrid threat modeling. While the bottom background layer groups the deterministic variables, the top layer represents the probabilistic variables. Probabilistic relations are displayed by black arrows and deterministic relations by grey arrows.

ence diagram by modeling both agents as players within a game-theoretical framework, allowing for the analysis of game equilibria.

Optimizing counter-hybrid strategy

A Bayesian network modeling technique is often used to account for the combination of probabilistic and deterministic relationships [24, 268, 260, 123, 256]. The truncated factorization of Chapter 3 is applied to factorize the joint probability distribution efficiently. Furthermore, the Bayesian networks will be extended to causal influence diagrams as in Chapter 4 to further dissect the nodes of the graph into random variables, utility nodes, and decision nodes and allow for causal interventions. In the first proposed modeling approach, the defender preemptively commits to a selected counter-hybrid posture to deter or dissuade agent A from conducting a hybrid operation, represented by a decision node. The adversary's response is driven by the estimated probability of successful dissuasion.

More formally, let D_D be the decision variable describing the counter-hybrid measure, which is followed by X_C, the random variable for the hybrid conduct. The available strategies and state spaces for D_D and X_C are $\Omega_{D_D} = \{d_1, \ldots, d_n\}$ and $\Omega_{X_C} = \{c_1, \ldots, c_m\}$ respectively. Each combination of counter-hybrid measure and hybrid operation variables leads to a discrete pay-off random variable X_Θ with state space $\Omega_{X_\Theta} = \{\theta_1, \ldots, \theta_k\}$ representing the interaction pay-off structure excluding direct costs. Because the pay-off structure from hybrid and counter-hybrid interaction are separated from the direct cost of hybrid operation and counter-hybrid measure, it is assumed that the pay-off structure has a zero-sum game component, meaning a positive value $\theta \in \Omega_{X_\Theta}$ corresponds to the gain of agent B and the loss of agent A, while a negative value for $\theta \in \Omega_{X_\Theta}$ represents a gain for agent A and the loss for agent B. The direct costs for the counter-hybrid measure and the hybrid attack are drawn from a cost probability function and denoted by U_Γ and U_H respectively, where the costs for the counter-hybrid measure has state space $\Omega_{U_\Gamma} = \{\gamma_1, \ldots, \gamma_n\}$ with $\gamma_i \geq 0$ for $i = 1, \ldots, n$ and the costs for the hybrid attack has state space $\Omega_{U_H} = \{\eta_1, \ldots, \eta_m\}$ with $\eta_i \geq 0$ for $i = 1, \ldots, m$. The total pay-off can finally be calculated by $U_{DP} = X_\Theta - U_\Gamma$ for agent B and $U_{CP} = -X_\Theta - U_H$ for agent A. The causal influence diagram that corresponds with these relations is depicted in Figure 6.5.[1]

[1]While the cost variables are priorly drawn from probability distributions, they become deterministic in the influence diagram, as only one cost value corresponds to a (counter-) hybrid measure per experiment.

Note that factorizing the direct costs in the initial pay-off node could have made the influence diagram purely probabilistic. However, for clarification purposes, direct costs have been separated from interaction costs. All probability distributions in the influence diagram are typically assumed to be categorical, which makes the conditional probabilities expressable via conditional probability tables.

Suppose that agent B has access to the potential costs of counter-hybrid measures $\Omega_{U_\Gamma} = \{\gamma_1, \ldots, \gamma_n\}$, their deterrence capacity $P(c \mid d)$ and their ability to mitigate potential damages $P(\theta \mid d, c)$. Agent B's objective is to compute the counter-hybrid measure that maximizes its total pay-off under the assumed probability distributions [9]. Specifically, the goal is to find the intervention $do(D = d)$ that maximizes agent B's expected total pay-off. Since the counter-hybrid measure node has no incoming arrows (see Figure 6.5), intervening on a variable is the same as conditioning on this variable [178]:

$$\begin{aligned}
&\max_{d_1,\ldots,d_n} \mathbb{E}[dp \mid do(D_D = d)] \\
&= \max_{d_1,\ldots,d_n} \sum_c \sum_\theta \sum_\gamma dp P(c, \theta, \gamma, dp, \mid do(D_D = d)) \\
&= \max_{d_1,\ldots,d_n} \sum_c \sum_\theta \sum_\gamma dp P(c \mid d) P(\theta \mid d, c) P(\gamma \mid d) P(dp \mid \gamma, \theta).
\end{aligned}$$

This can be formulated as an integer linear program (ILP):

$$\begin{aligned}
&\max \sum_{h=1}^{k} \sum_{j=1}^{m} \sum_{i=1}^{n} \theta_h w_{ijh} q_{ij} p_i + \sum_{i=1}^{n} \gamma_i p_i \\
\text{subject to} \quad &\sum_{i=1}^{n} p_i = 1 \\
&p_i \in \{0, 1\} \quad i = 1, \ldots, n
\end{aligned}$$

Subgame perfect equilibrium

In optimizing the counter-hybrid strategy, probabilities are used to estimate the likelihood of successfully deterring the adversary after each measure. These probabilities are determined ex-ante, meaning they are drawn before a counter-hybrid measure is chosen. As a result, they do not account for any short-term pay-off adjustments that occur during the interaction that eventually set the outcome in equilibrium. To this end, the causal influence diagram approach is extended to multi-agent influence diagrams [128, 90] to account for these strategic considerations, and the notion of

equilibrium is addressed using the causal games that emerge from such diagrams.

Formally, the hybrid conduct node X_C of Figure 6.5 is no longer modeled probabilistically but is instead represented as a decision node D_C. Additionally, the distinction between the two agents, A and B, is made explicit by assigning decision and utility nodes to each respective player. Specifically, decision node D_D and utility nodes U_{DP} and U_Γ are assigned to agent B, while decision node D_C and utility nodes U_{CP} and U_H are associated with agent A.

A solution concept is necessary that pinpoints a subset of possible outcomes when agents act rationally. As explained in Chapter 4, the subgame perfect equilibrium is a natural solution concept in the causal games that emerge from MAIDS, which helps eliminate non-credible threats. In the context of the hybrid threat game, non-credible threat equilibria emerge when the attacker threatens to conduct a hybrid operation despite it not being in their best interest in terms of pay-off.

Probability distributions and elicitation

Filling the influence diagram with accurate conditional probabilities is widely recognized as a challenging task [123] and a rigorous elicitation process should be developed to ensure the highest degree of accuracy in the inputs. To maintain a realistic perspective in the estimates, the cost, potential deterrence capacity, (either denial or punishment), and resilience-enhancing ability of each counter-hybrid measure are estimated on the basis of an in-depth literature review complemented with a mini-Delphi approach with seven (junior) analysts with a background in strategic studies. This resulted in probability distributions from which samples were drawn to conduct the experiments. Initially, the parameters of these distributions were inspired by a literature review. Subsequently, analysts made a one-time adjustment to the parameters, informed by visualizations of the resulting distributions. The specifics of these probability distributions per variable are summarized in Table 6.3 while the exact parameters are available in Appendix C.1. Values that are likely to be drawn from these probability distributions indicate that they align closely with consensus in the literature and the outcomes of the mini-Delphi survey, while values unlikely to be drawn correspond to values that are less in alignment. By repeatedly sampling input variables from these distributions independently, a total of 1000 experimental scenarios were generated. These experiments can be considered semi-synthetic due to the absence of a rigorous, standardized method for constructing the prior distributions for these estimates, as described in Section 4.2.4, requiring reliance on the constructed probabilistic representations.

Table 6.3: Values drawn from probability distributions. Costs of counter-hybrid measures and damaging impacts of hybrid attacks are expressed in US million dollars. While the costs of counter-hybrid measures and damaging impacts of hybrid attacks are drawn from variants of the normal distribution, the probability values for the ability to deter and the ability to mitigate damaging impacts are drawn from their corresponding conjugate priors (Beta and Dirichlet, respectively).

Value	Meaning	Probability Distribution
θ_h	Potential damage of a hybrid attack	Drawn from different truncated normal distributions for each category of damaging impacts [170]
γ_i	Cost for conducting counter-hybrid measure d_i	Drawn from different truncated normal distributions for each counter-hybrid measure d_i [170]
q_{ij}	Probability of the adversary conducting hybrid operation c_j after counter-hybrid measure d_i	Drawn from different Beta distributions for each counter-hybrid measure
w_{ijh}	Probability of potential damage θ_h after the adversary conducts hybrid conduct c_j and defender counter-hybrid measure d_i	Drawn from different Dirichlet distributions for each counter-hybrid measure and hybrid operation combination [228]

Despite the semi-synthetic nature of the experiments, all the counter-hybrid measures considered have been derived from real-world examples and their impacts have been scored by experts, ensuring reflection of real-world variability and available empirical evidence. As the modeling approach enables the exploration of dynamics that cannot be empirically tested in the real world, the semi-synthetic nature of the data is a necessary instrument to conduct this analysis. Furthermore, the flexibility of the proposed framework ensures its applicability to other domains and hybrid threat types, as the underlying principles and interactions are generalizable beyond the specific scenarios tested. This adaptability enhances its utility in addressing a broad spectrum of hybrid threat challenges.

6.2.2 Experimental Design

A scenario is considered in which the defending agent B fears that revisionist agent A attempts to destabilize and harm agent B through hybrid attacks. In particular, the defender is aware of agent A's offensive capabilities in the cyber and information domains and is concerned that the latter will carry out a high-scale cyber-attack against its critical infrastructures, such as power plants, and grids, water management facilities, ports, the healthcare system and/or other essential services. Offensive cyber

operations constitute a clear example of a hybrid threat below the threshold of large-scale armed conflict. Indeed, cyber operations have become more prevalent in recent years due to the technical, physical and logical layers of cyberspace and the pervasive use of networks and technologies in our daily life [14].[2] Furthermore, offensive cyber operations may well produce material consequences resulting in considerable physical damage such as for instance in the case of Stuxnet in Iran (2009), Shamoon in Saudi Arabia (2012) or NotPetya in over sixty countries around the world (2017).

An anonymized list of plausible hybrid actions was constructed based on a series of real-world malicious cyber operations drawn from the updated datasets compiled by Valeriano and Maness [243], Stirparo et al. [234], and the Council on Foreign Relations,[2] as well as on a review of the relevant literature. In addition, given the exponential development of new technologies and the evolving dynamics in current conflicts, this was complemented with expert imagination, in an effort to anticipate potential courses of action (and response), and key variables in the cyber domain were distilled. Plausible counter-hybrid responses were identified, with experts selecting the top five cross-domain measures against malicious cyber attacks, summarized in Table 6.4. These include both in-domain (cyberspace) and out-domain responses, such as law enforcement, norm development, public diplomacy, and economic sanctions. Viewed through cumulative deterrence, some measures aim to mitigate damage, while others seek to deter aggression by altering adversaries' cost-benefit assessments. Alternatively, the defender can choose to refrain from engaging in counter-hybrid measures. Appendix C.1 contains the specifications of the measures, including probability distributions that will be used to distill the costs, the probability of successful deterrence, and the probability of mitigating damaging effects for each of the different counter-cyber measures.

Estimating the exact costs and damages from cyber attacks on critical infrastructure is challenging [88], but experts agree that the defender's capacity to detect and recover from such attacks significantly shapes their overall impact [163, 237]. Accordingly, three categories of impact are considered: θ_1 denotes severe disruption due to ineffective detection and recovery; θ_2 reflects substantial losses with partial or delayed mitigation; and θ_3 captures limited or negligible effects resulting from effective preventive or responsive measures. As affected entities rarely disclose precise impact data, potential impacts θ_1, θ_2, and θ_3 are sampled from heavy-tailed half-normal distributions, following Lis and Mendel [146], with specifications provided in Appendix C.1 to

[2]Council of Foreign Relations, "Cyber Operations Tracker," accessed December 1, 2022, `https://www.cfr.org/cyber-operations/`

Table 6.4: Five Counter-Hybrid Measures

Domain	Title	Capability	Rationale
Military	Active intelligence sharing	Actively share your collective intelligence with allies.	Active intelligence sharing is useful for bringing your allies on board for political, public or other types of collective attribution.
Cyber	Boost cyber resilience at the wider societal level	Introduce legislation or collaboration that requires individuals and companies to adopt sufficient levels of cyber resilience, based on the specific risk exposure of the subject.	Boosting cyber resilience at the broader societal level is one of the core tenets of a whole-of-society approach to hybrid/cyber threats.
Cyber	Employ offensive cyber capabilities	Use offensive cyber operations in order to undermine the target.	It is often the case that cyber attacks are not directly attributable. In the short term, it provides a covert way to influence the target. The effect of an offensive cyber attack is scalable.
Legal	Market restrictions	Introduce legislation to restrict an opponent from accessing your market in a specific sector (such as ICT).	Such a measure reduces the possibilities for an adversary to exploit vulnerabilities but it may also clash with other legal commitments (see international trade commitments against market restrictions based on nationality).
Diplomatic	Open deterrence messaging through strategic communications	Communicate one's strategic posture in order to convince a target to comply with one's strategic aims.	Being transparent with the hostile actor regarding one's own strategic strengths and possible actions. This increases the possibility of a better-informed decision by the hostile actor.

capture the high variance and uncertainty inherent to such estimates.

6.2.3 Results

In this section, the results are discussed based on the experimental setup of the previous section. First, the results of optimizing for the counter-hybrid measure are presented

using estimated deterrence probabilities. This is followed by an analysis of the subgame perfect equilibria, where the decision to conduct the hybrid attack is modeled as the agent's strategic choice.[3]

For each of the 1000 experiments, the effectiveness of the counter-hybrid measure is ranked in terms of total pay-off for defending agent B from least optimal to most optimal. A count plot of the rank of the counter-hybrid measures for all experiments is displayed in Figure 6.6.

In summary, despite the high cost of imposing market restrictions (d_4), they are often deemed the most optimal counter-hybrid measure due to the potential to mitigate damages and to deny the adversary's ability to carry out attacks. Intelligence sharing (d_1), valued for its cost-effectiveness, plays a crucial role in mitigating attack damage by enabling timely defensive actions and fostering political support for collective responses. While offensive cyber operations (d_3) could disrupt enemy capabilities, they carry high risks of escalation and have variable success rates. Boosting cyber resilience (d_2), though costly, is consistently rated effective for both damage mitigation and deterrence. Open deterrence communication (d_5) hinges on the adversary's responsiveness to threats, requiring detection, attribution, and communication capabilities to be successful. Lastly, in very rare draws, abstaining from counter-hybrid

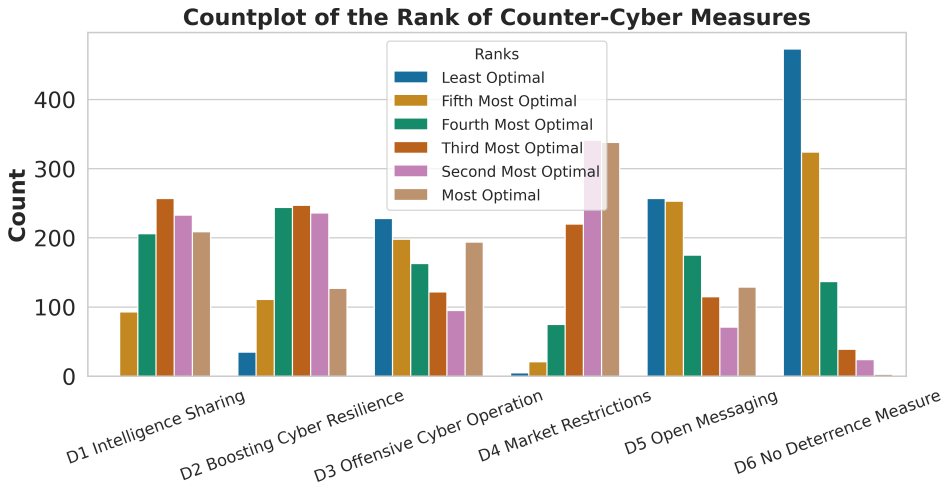


Figure 6.6: The count of the specific rank that each of the counter-hybrid measures is computed to attain.

[3]The modeling effort is publicly available at `https://github.com/HCSS-Data-Lab/Hybrid-Threat-Implementation`.

measures (d_6) emerges as the most effective strategy.

To derive more meaningful insights, the focus is shifted from the specific outcomes of individual measures to the broader results that can be drawn from the overarching characteristics of these measures, especially considering that a well-designed elicitation protocol would significantly enhance the interest and reliability of individual results, while the same overarching characteristics would prevail.

Overall, the measures vary in several ways: some measures rely on their ability to dissuade the adversary through punishment (open messaging, offensive cyber operations), others count on the ability to mitigate the potential damage when confronted with an attack (intelligence sharing), and there are also measures that are a mixture of both deterrence by denial and enhancing resilience (boosting cyber resilience and imposing market restrictions). While these characteristics are distributed evenly among the optimal measures, the measure designated as optimal in most cases - i.e., imposing market restrictions - is also the most versatile one with respect to both dissuasion as well as resilience enhancement. In addition, the variance of the cost, ability to mitigate the damage and ability to deter are different for each of the measures as their impact is mediated by favoring conditions. For instance, while deterrence by punishment measures (open messaging and offensive cyber operation) can be very effective counter-hybrid measures, they are also among the most ineffective measures for some experiments as illustrated by Figure 6.6. This is because they rely heavily on their effect on the adversary's strategic calculus. When dissuasion is not successful, they do not contribute to mitigating the damaging impact of hybrid conduct, leaving the defender exposed.

In order to test how variations in input parameters influence the output of the model, sensitivity analyses were conducted using the state-of-the-art tool of van Stein et al. [233] for each of the counter-hybrid measures. Figure 6.7 presents the SHAP summary plot for imposing market restrictions, which serves as a representative example of the broader sensitivity analysis conducted. As can be observed from the figure, the probability of successfully deterring the adversary q_{14}, as well as the costs of the measure γ_4 are the most influential factors for the effectiveness of the counter-hybrid measure. While the measure's ability to mitigate the negative impact is comparatively less influential overall, its role in increasing the probability of a negligible impact (ω_{413}) or reducing the likelihood of a severe impact (ω_{411}) remains a significantly important factor in evaluating its effectiveness. As the effectiveness of the measure is strongly influenced by its ability to dissuade the adversary, there is a need to consider this factor not only probabilistically but also within a game-theoretic framework. By doing

so, the subgame perfect equilibria were computed as detailed in the previous section, providing a more comprehensive evaluation of the measures in a strategic context.

These subgame perfect equilibria for the same 1000 experiments are displayed in Table 6.5, reflecting outcomes where agents seek to optimize their pay-off rationally. The low occurrence of hybrid attacks indicates that the chosen counter-hybrid measures focus on deterring the adversary rather than mitigating the consequences of a hybrid attack. Given that the adversary's strategic calculus is assumed to be known, the defending agent can strategically select the most cost-effective counter-hybrid measure that successfully deters the adversary from launching an attack. This explains why intelligence sharing is preferred over market restrictions when both measures suffice to deter the adversary. Moreover, when the strategic calculus is such that the

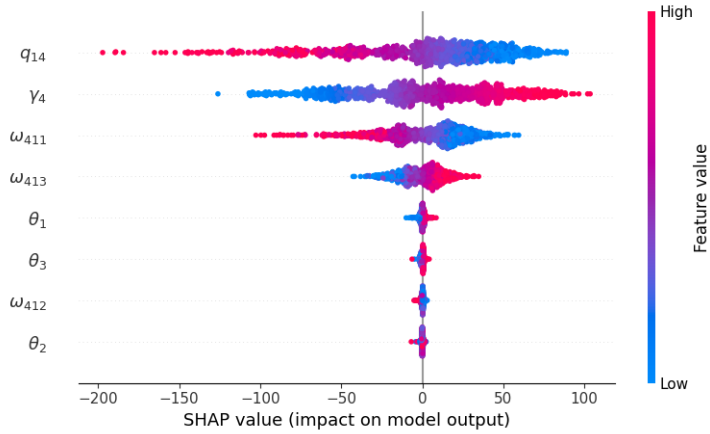


Figure 6.7: SHAP summary plot for counter-hybrid measure imposing market restrictions: The y-axis represents the features ranked by their importance to the model output. The x-axis shows the SHAP value, indicating the magnitude and direction of each feature's impact on the model output. The color gradient reflects the feature values.

Table 6.5: Subgame Perfect Equilibria Outcome Occurences

Counter-hybrid Measures	Hybrid Operation	
	Attack	No Attack
D1 Intelligence Sharing	15	697
D2 Boosting Cyber Resilience	1	98
D3 Offensive Cyber Operation	3	76
D4 Market Restrictions	0	110
D5 Open Messaging	0	0
D6 No Deterrence Measure	0	0

adversary is likely to proceed with a hybrid operation regardless of the counter-hybrid measure, the subgame perfect equilibria suggest that the defending agent should commit to cost-efficient counter-hybrid measures, such as intelligence sharing, to mitigate the impact of the attack.

6.2.4 Conclusion and Future Work

This section proposed novel approaches to evaluate cross-domain counter-hybrid measures by balancing cost, deterrence, and damage mitigation under uncertainty, modeled probabilistically and game-theoretically. The evaluation included 1000 scenarios of malicious cyber operations, incorporating countermeasures derived from literature and a mini-Delphi survey.

While general validation of the results remains challenging due to the ambiguous nature of hybrid threats and the reluctance of targeted parties to disclose information, contextualizing the findings within established models and prior studies provides valuable insights. The results confirm that the most effective counter-hybrid measures tend to be the most costly, supporting previous claims that hybrid threats can economically strain defenders [25], and underscoring the importance of prioritizing cost-effective, cross-domain strategies [18]. The analysis also supports democratic deterrence models advocating a whole-of-society approach [242], particularly highlighting the utility of non-military measures like market restrictions and intelligence sharing [258]. Finally, the sensitivity analysis reinforces the findings on existing modeling efforts on deterrence in the cyber realm [130], emphasizing that the effectiveness of such threats is heavily contingent upon the adversary's susceptibility to countermeasures. Deterrence by punishment measures, for instance, only work well when the adversary is responsive to such measures [158]. The modeling exercise indicates that even a small enhancement in understanding the aggressor's plausible receptiveness to counter-hybrid measures could lead to a significant enhancement in the assessment of the effectiveness of measures. This implies that resources spent on anticipating the adversary's reaction to possible counter-hybrid measures are conditional on the effect of the counter-hybrid measures.

Two key directions for future research are identified. First, applying elicitation methods for probabilities and utilities from Chapters 3 and 4, in direct consultation with policy-makers, would improve the precision of model inputs. Second, extending experiments to cross-domain hybrid threat scenarios beyond the cyber domain would increase realism by better capturing the complexity of modern hybrid threats.

Chapter 7

Conclusions and Future Work

This thesis advances the application of (strategic) causal frameworks within the context of complex security environments. It does so by surveying and developing methodological approaches that are subsequently applied to substantive policy challenges. This final chapter provides a summary of all previous chapters and thereby discusses how this thesis contributes to addressing the research questions. The conclusion then turns to address the broader implications and contributions of this research, structured around the challenges of causal inference, its policy relevance, remaining research gaps, and the thesis's theoretical, methodological, empirical, and societal significance. Finally, future research avenues are proposed.

7.1 Summary

Chapter 1: The introduction establishes the context for integrating causal inference with strategic interaction in complex security environments. It identifies a critical research gap: while game theory is widely used in strategic studies, causal inference remains notably underutilized despite its importance for evidence-based policy-making. This gap persists because complex security environments—featuring interference effects, strategic interdependence, and adaptive adversaries—violate fundamental assumptions of standard causal models. Additionally, the mathematical complexity and computational demands of existing causal methods create implementation barriers for practitioners. The chapter outlines how this dissertation addresses these challenges through an integrated framework that enables the application of causal inference in complex security environments. The research questions are structured to examine the

necessary conceptual frameworks (RQ1), develop computationally efficient methods (RQ2), and demonstrate practical applications (RQ3) in such security environments.

Chapter 2: The preliminary chapter presents the fundamentals of probability theory, graphical models, and game theory. These elements establish the foundational terminology and conceptual framework that underpin the development of the subsequent chapters.

Chapter 3: This chapter organizes fragmented causal inference methods into Pearl's causal hierarchy. The hierarchy contains three levels: association (seeing), intervention (doing), and counterfactual (imagining). Each level requires specific mathematical tools. The associational level uses Bayesian networks to model observational relationships. The interventional level employs causal Bayesian networks and do-calculus to estimate causal effects. The counterfactual level needs structural causal models for retrospective analysis. By identifying which causal concepts belong to each level, this chapter answers RQ1.1: *What fundamental causal concepts are necessary for structuring and differentiating causal relationships, particularly in the context of Pearl's causal hierarchy?* Additionally, the chapter reveals critical assumptions for each method, answering RQ1.2: *What key assumptions underpin causal inference applications across Pearl's causal hierarchy?* It shows how relaxing standard assumptions necessitates alternative causal structures that explicitly model the resulting complexities. It also distinguishes parametric methods from non-parametric methods. For contexts where the no-interference assumption fails, the chapter adopts the spatially explicit structural equation model as an alternative framework that explicitly models spillover effects. The chapter equips practitioners with decision criteria to select methods based on their policy questions.

Chapter 4: This chapter integrates causal reasoning with strategic decision-making to model complex security environments. The chapter introduces game-theoretic foundations (normal-form, extensive-form, and Bayesian games) before extending them to incorporate causal structures. Multi-agent influence diagrams combine game theory with causal graphs, enabling simultaneous modeling of strategic interactions and causal mechanisms. These form the basis of causal games, where equilibrium strategies induce causal Bayesian networks. By showing how causal games enable policy-makers to compute both strategic equilibria and causal intervention effects within a unified framework, the chapter answers RQ1.3: *What methods exist for integrating causal rea-*

soning with strategic decision-making in complex security environments, and how can they be applied? The chapter provides practical guidance on model selection based on problem structure and details the elicitation requirements for implementation. Examples from deterrence scenarios demonstrate how this framework captures the interplay between causal effects and strategic adaptation in complex security environments.

Chapter 5: The most technical contributions of this thesis are presented in this chapter, where a method is presented to approximate optimal causal interventions in hybrid Bayesian networks. The chapter first introduces an approach for approximate inference based on discretization and decision diagrams. This approach achieves over 10x speedup compared to traditional methods, while Pareto fronts reveal how practitioners can balance computational cost against inference accuracy. These visualizations show precisely how many discretization bins to use for desired accuracy levels, providing implementation guidance. This systematic analysis of the accuracy-efficiency trade-off answers RQ2.1: *How can inference be performed efficiently to accurately estimate the effects of causal interventions while maintaining computational feasibility?* Finally, this method is then embedded in a broader framework that, in combination with optimization algorithms, can approximate optimal interventions in such hybrid Bayesian networks. Empirical evidence shows that while local methods (BFGS, Powell, OnePlusOne) perform poorly, differential evolution and NGOpt consistently outperform random search, demonstrating that these problems contain sufficient structure for heuristic optimizers to exploit. This performance comparison across multiple optimization techniques answers RQ2.2: *How can optimization techniques be integrated with causal inference to optimize over causal interventions efficiently under budget constraints?*

Chapter 6: This chapter applies the previously introduced concepts and validates the scientific contributions in the context of complex security environments, thereby addressing RQ3: *How can the proposed (strategic) causal concepts be applied to complex security environments?* Two applications are examined. First, causal frameworks from Chapter 3 are applied to environmental conflict in Iraq. A causal discovery algorithm uncovers the empirical structure linking environmental factors to conflict from 294 municipal observations. The analysis reveals that soil moisture and latent energy affect conflict through demographic and agricultural mediators. Spatially explicit structural equation models estimate these effects while accounting for interference between municipalities. This application proves that causal methods can identify

actionable intervention points in conflict prevention. The second application examines hybrid threat deterrence using the causal game-theoretic framework from Chapter 4. A causal influence diagram models how states select counter-hybrid measures against cyber attacks on critical infrastructure. Integer linear programming identifies market restrictions and intelligence sharing as optimal counter-hybrid measures across 1000 scenarios. The multi-agent extension computes subgame perfect equilibria, showing that the deterrer prioritizes cost-effective deterrence over damage mitigation when adversaries act strategically. Sensitivity analysis confirms that the effectiveness of counter-hybrid measures depends primarily on adversary responsiveness. This application demonstrates how causal game theory informs security policy under strategic contestation.

7.2 Conclusions

Causal inference has evolved significantly across disciplines since David Hume's 18th-century philosophical inquiries into the nature of causation. Medicine advanced the field through the development of randomized controlled trials in the mid-20th century, establishing rigorous methods for determining treatment effects. Computer science has recently transformed causal inference from observational data, with Judea Pearl developing graphical models and the do-calculus for causal reasoning [178], Peter Spirtes pioneering constraint-based algorithms for causal discovery [231], and Elias Bareinboim advancing methods for causal inference across heterogeneous domains [28].

While these sophisticated causal methods have remained largely confined to computer science and clinical domains, strategic studies would benefit immensely from their systematic application to evaluate policy interventions. Policy-makers must distinguish between actions that genuinely reduce conflict and those that merely correlate with peaceful periods. Without causal analysis, security interventions risk squandering resources on ineffective measures or inadvertently escalating conflicts through misguided policies. Accurate prediction of causal intervention outcomes becomes particularly essential in complex security environments. In these environments, interventions trigger cascading effects that spread across regions and organizations, making it crucial for policymakers to understand both immediate outcomes and secondary consequences. Moreover, adversaries actively exploit gaps in causal understanding by adapting their tactics to circumvent interventions, shifting operations to areas where policy effects are weakest.

However, applying causal inference to complex security environments faces signifi-

cant obstacles. Standard causal models rely on assumptions that these environments often violate, especially the no-interference assumption, which requires that the intervention of one unit does not affect the outcome of another. In practice, policy interventions can cause armed groups to move to neighbouring areas, creating spillover effects that undermine these assumptions. Additionally, traditional causal frameworks struggle to model multiple strategic actors who anticipate and adapt to interventions. Adversaries observe security policies and change their tactics, altering the very causal relationships that analysts aim to understand. These technical limitations, along with the mathematical complexity of advanced causal methods, create a considerable implementation gap, preventing security practitioners from using tools that could improve their decision-making.

This dissertation addresses these challenges by extending causal inference beyond its traditional computational foundations, adapting its core concepts to be more applicable within complex security studies. By clarifying foundational concepts in both causality and game theory, it demonstrates how these analytical tools can be integrated to tackle the unique problems of complex security environments. The thesis illustrates how causal reasoning, when combined with models of strategic interaction, provides a robust framework for understanding and guiding policy interventions in environments defined by uncertainty, interdependence, and competing objectives.

The findings of this research reveal that concepts from causal inference, when properly adapted to account for the characteristics of complex security environments such as interference and strategic interactions, provide a powerful framework for addressing these environments. Specifically, the research demonstrates that while sophisticated causal methods already exist in other domains, such as spatially explicit structural equation models in ecology, they require systematic extraction, restructuring, and adaptation to become operational in strategic studies. By organizing these disparate methods within Pearl's causal hierarchy and connecting them to specific policy questions, this thesis establishes how causal tools can be systematically applied to improve decision-making in security environments characterized by uncertainty, interdependence, and competing interests.

The thesis further concludes that the implementation gap between theoretical causal methods and practical security applications can be bridged through computational innovation. The development of efficient approximation methods for optimal causal interventions in hybrid Bayesian networks demonstrates such innovation. By achieving over 10x speedup compared to traditional methods while maintaining accuracy, and by providing clear visualization of accuracy-efficiency trade-offs, these tools

enable policy-makers to make informed choices about computational resources versus analytical precision. This computational contribution transforms abstract causal theory into actionable decision support.

Finally, this thesis concludes that causal game-theoretic frameworks generate actionable policy insights when applied to real-world security contexts. The case studies on environmental conflict in Iraq and hybrid threat deterrence demonstrate that these models can effectively handle the defining features of complex security environments: spatial interference between units, strategic adaptation by multiple interdependent actors, and pervasive uncertainty. The empirical findings reveal that latent energy and soil moisture indirectly cause conflict activity through demographic and agricultural mediators, while the game-theoretic analysis identifies which characteristics of counter-hybrid measures effectively deter adversaries from conducting hybrid operations. These results demonstrate that this integrated approach enhances security policy effectiveness by anticipating strategic responses and enabling targeted interventions based on causal understanding.

In conclusion, this research provides a structured, accessible approach to the application of causal inference within strategic studies. The contributions of this thesis span multiple dimensions, each addressing distinct aspects of the challenge of applying causal inference to complex security environments. The following sections detail how this work advances the field through theoretical, methodological, empirical, and societal contributions.

Theoretical Contributions: The dissertation presents a structured framework for integrating causal inference and strategic interaction in complex security environments. Pearl's causal hierarchy, which organizes causal reasoning across three levels (association for observational relationships, intervention for effects of manipulations, and counterfactuals for retrospective analysis of alternative scenarios), provides the foundational structure. The first theoretical contribution of this dissertation systematically maps existing causal concepts and their underlying assumptions onto this hierarchy, providing practitioners with clear guidance on which causal tools are appropriate for specific security policy questions. This mapping enables more effective alignment between policy questions and analytical methods by making explicit the assumptions required at each level.

Recent work has begun integrating game-theoretic elements into causal models [91], but these approaches have remained largely bereft of real-world application due to their technical complexity. The second theoretical contribution enhances the practical accessibility of these existing methods by explicitly specifying the foundational game-

theoretic elements, clarifying their intersection with causal concepts, and detailing the model inputs and data requirements needed for implementation in security contexts. This operationalization bridges the gap between theoretical frameworks and practical application.

Methodological Contributions: To address implementation barriers, the thesis develops computational innovations that make sophisticated analytical tools practically viable for security practitioners. The primary methodological contribution is a novel approach for approximating causal interventions in hybrid Bayesian networks through discretization and knowledge compilation. This method directly addresses the computational constraints that can prevent practitioners from applying analytical tools. It demonstrates through empirical evaluation how the approach balances inference accuracy with computational feasibility. The trade-off between computational cost and accuracy is systematically analyzed and visualized through Pareto fronts, which provide practitioners with guidance on parameter selection based on their specific accuracy and resource requirements. The dissertation embeds the approximation method within an optimization framework that enables evaluation of multiple causal interventions. This framework benchmarks various optimization algorithms and allows practitioners to identify optimal interventions while respecting real-world resource limitations.

Empirical Contributions: This dissertation demonstrates practical applicability through two applications that exemplify complex security challenges. The first application analyzes environmental conflict in Iraq, applying causal discovery methods to uncover mechanisms linking environmental factors to violence. The analysis uses literature-informed variable selection to guide causal discovery and accounts for spatial interdependencies in computing causal estimates of conflict incidence across geographical units.

The second application focuses on hybrid threat deterrence, demonstrating how strategic causal models can inform counter-hybrid strategies. This case models adversarial hybrid operations using causal influence diagrams and uses integer linear programming to identify optimal counter-hybrid measures. The application extends to multi-agent settings where subgame perfect equilibria are computed to determine optimal strategies under strategic contestation, with sensitivity analysis providing insights into how different variables impact equilibrium outcomes.

Societal Contributions: These empirical applications demonstrate the thesis's broader societal and policy contribution by validating how strategic causal models inform real-world decision-making. The Iraq case study reveals how environmental

variables drive conflict through specific causal pathways that enable targeted policy responses that account for spatial spillovers between regions. The hybrid threat analysis quantifies the effectiveness of different deterrence measures, helping policymakers optimize resource allocation across defensive measures. These findings illustrate the practical value of integrating causal reasoning with strategic elements to support evidence-based policy in contested environments. Beyond security contexts, this approach applies to any domain where policymakers must navigate both causal complexity and strategic interdependence among competing actors.

While this thesis advances the integration of causal inference and strategic interaction, significant challenges remain that limit the full realization of these methods in complex security environments. The contributions presented here represent important steps forward, yet they also illuminate critical gaps that must be addressed before causal game-theoretic frameworks can reach their full potential in such security environments.

Theoretically, Pearl's causal hierarchy rests upon structural causal models where recursiveness is often assumed. However, most complex security environments exhibit inherent feedback loops and cyclical dynamics. Escalatory dynamics, where actions and counteractions spiral through cycles of increasing intensity, are fundamentally cyclic in nature and violate standard recursiveness assumptions. The theoretical foundations for integrating cyclic causal models with strategic interaction remain underdeveloped, limiting the framework's ability to capture these dynamics in strategic settings.

Methodologically, the no-interference assumption, which requires that interventions for one unit do not affect others' outcomes, is highly unlikely to hold in complex security environments where spillover effects are ubiquitous. While this thesis introduces spatially explicit structural equation models to partially address this limitation, the vast majority of causal research continues to rely on no-interference assumptions. Methods that can simultaneously handle various forms of interference, while accounting for strategic interaction, remain computationally intractable or theoretically incomplete.

Empirically, real-world security settings often suffer from limited, sensitive, and inconsistently structured data due to secrecy, classification, and covert activities. At the same time, there is a lack of rigorous empirical validation, as it is rarely possible to evaluate the effectiveness of real-world policy interventions. Most causal game-theoretic methods require complete data or rely on strong assumptions when data is missing, making them ill-suited to these challenges. Strategic concealment further

complicates both data availability and validation, reinforcing the gap between theory and practice.

7.3 Future Work

Numerous avenues for advancing the theory, methodology, and empirical validation have been explored throughout this thesis. The present discussion is limited to future directions aimed specifically at addressing the broader research gaps outlined above.

Addressing the theoretical gaps requires extending causal frameworks beyond their current acyclicity constraints. Research is now emerging that accounts for the existence of feedback loops in structural causal models [36], but more research is necessary to fully integrate these cyclic structures with strategic contestation. Future work should develop mathematical foundations that can represent escalatory dynamics and feedback loops while maintaining the analytical tractability needed for policy applications. This includes establishing identification criteria for causal effects in cyclic strategic systems and developing equilibrium concepts that account for how causal relationships evolve through repeated strategic interactions.

To overcome methodological limitations, future research must develop scalable approaches for relaxing the no-interference assumption in strategic settings. Computational tractability is particularly critical here, as approximating causal interventions under the no-interference assumption is already computationally demanding, and introducing interference effects multiplies these challenges significantly. This highlights the need for targeted research into which specific forms of interference are most relevant in complex security environments, followed by the development of specialized interference models that maintain computational tractability as a primary design constraint. Promising directions include leveraging machine learning advances to approximate complex interference patterns efficiently and creating diagnostic tools that help practitioners systematically identify which types of interference are most critical in their specific contexts.

With respect to empirical data constraints, tailored data collection strategies must be developed to accommodate the specific characteristics of complex security environments. For example, in environmental security, international efforts to compile climate-conflict indicators have enabled the environmental conflict analysis in this thesis. Similarly, advancing causal models in domains such as hybrid threat deterrence will require investment in datasets capturing cyber activities, influence operations, and multi-actor strategic behaviour. For domains where enhanced data availability

remains unlikely, greater emphasis should be placed on structured expert elicitation, including formal extraction of utilities, belief distributions, and strategic type assessments from domain experts. While methods exist for discrete settings, further work is needed to adapt these techniques for continuous domains and complex interdependencies. Regarding validation constraints, monitoring mechanisms must be built to assess the effectiveness of interventions in real-world security contexts. This should enable systematic evaluation of whether theoretical predictions translate into practical outcomes and allow for further refinement of causal game-theoretic frameworks.

Finally, a valuable direction for future research lies in the development of plug-and-play tools that enable users to apply causal game-theoretic analysis without requiring proficiency in programming or advanced formal modeling. Such tools should offer predefined templates, guided workflows, and user-friendly interfaces that allow practitioners to input domain-specific knowledge and data, while automating the underlying computational procedures. Embedding these methods within the curricula of strategic and security studies, particularly as part of quantitative methods training, would further support their diffusion. Doing so would equip future analysts and decision-makers with accessible tools for causal reasoning in strategic contexts, without placing undue technical demands on users.

Appendix A

Acronyms and Notation Conventions

A.1 Acronyms

A unified list of acronyms that are frequently used throughout the thesis is presented in this section. It includes terms frequently encountered in the context of causal game theory, as well as those related to the methodology introduced in Chapter 5, such as discretization, decision diagrams, and optimization.

ADMG Acyclic Directed Mixed Graph
ATE Average Treatment Effect
BDD Binary Decision Diagram
BN Bayesian Network
BP Belief Propagation
CAIM Class-Attribute Interdependence Maximization
CBN Causal Bayesian Network
CM ChiMerge
CNF Conjunctive Normal Form
CPDAG Completed Partially Directed Acyclic Graph
CPT Conditional Probability Table
DAG Directed Acyclic Graph
DDN Dynamic Discretization
DE Differential Evolution

EBP Bayesian Method with Adjusted Empirical Bayes Priors
EF Equal Frequency
EFG Extensive-Form Game
EW Equal Width
FFRCISTGS Finest Fully Randomized Causally Interpretable Structured Tree Graph
GES Greedy Equivalent Search
ID Influence Diagram
LiNGAM Linear Non-Gaussian Acyclic Model
MAID Multi-Agent Influence Diagram
MDLP Minimum Description Length Principle
MLE Maximum Likelihood Estimation
NFG Normal-Form Game
NPSEM-ie Non-Parametric Structural Equation Model with Independent Errors
PE ATE Percentage Error of the Average Treatment Effect
PGM Probabilistic Graphical Model
PO Potential Outcome
POF Potential Outcome Framework
RCT Randomized Controlled Trial
SCM Structural Causal Model
SEM Structural Equation Model
SESEM Spatially Explicit Structural Equation Model
SPE Subgame Perfect Equilibrium
SUTVA Stable Unit-Treatment Value Assumption
SWIG Single World Intervention Graph
VE Variable Elimination
WRMSE Weighted Root Mean Squared Error

A.2 Notation Conventions

The notation conventions that are used throughout this thesis can be found in Table A.1.

Table A.1: List of Frequently Used Notation Conventions

Symbol	Object
$\boldsymbol{A}$	Set of Action Sets
B_j	Discretized Bin
Γ	Game
$\mathbf{ch}(V_i)$	Children of Node V_i
$\boldsymbol{D}$	Set of Decision Nodes
do	Do-Operator
$\mathbf{de}(V_i)$	Descendants of Node V_i
$\boldsymbol{E}$	Set of Edges
$\mathbb{E}[X_i]$	Expected Value of X_i
$\boldsymbol{F}$	Set of Structural Functions
$\mathcal{G}$	Family of Graphs
G	Graph
G_i	Subgraph of Nodes V_j, where $j \leq i$ in the Topological Sort
$G_{\overline{\boldsymbol{V}'}}$	Mutilated Graph G with respect to $\overline{\boldsymbol{V}'}$
M	Agents
μ_i	Higher-Order Belief
$\mathbf{nonde}(V_i)$	Non-Descendants of Node V_i
ρ	Path
$\mathbf{pa}(V_i)$	Parents of Node V_i as Random Variables
$\boldsymbol{pa}_i$	Assignments of Parents of Node V_i
$P(X_i)$, $p(X_i)$	Probability Distribution of X_i (Discrete or Continuous)
$\mathcal{S}$	Structural Causal Model
σ^i	Strategy
σ	Strategy Profile
σ^{-i}	Partial Strategy
T	Treatment Variable or Type Profile Tuple
$\boldsymbol{U}$	Set of Utilities
$\boldsymbol{V}$	Set of Nodes
$\boldsymbol{W}$	Set of Exogenous Variables
$\boldsymbol{X}$	Set of Random Variables
X_i	Random Variable
$\boldsymbol{X}_{-i}$	Random Variables $\boldsymbol{X} \setminus X_i$
$\boldsymbol{x}$, x_i	Assignment of $\boldsymbol{X}$, X_i
Y	Dependent or Outcome Variable
Ω	State Space
Ω_{X_i}	State Space of Random Variable X_i
$\perp\!\!\!\perp_P$	Independence in Probability
$\perp\!\!\!\perp_G$	d-separation
$<$	Ordering induced by Topological Sort

Appendix B

Appendix Chapter 5

B.1 Heatmap Results

As the computational cost becomes generally higher when the number of bins in the discretization process increases, the heatmaps in this section focus on the quality of discretization and inference. For example, the discretization EF9 in Pareto front of Figure 5.5e dominates all other EF discretizations in terms of the error. Therefore, EF9 returns in the heatmap of Table B.1 in the corresponding error and CPT inference method column.

Table B.1: Heatmap summarizing the causal results: every box refers to a Pareto front corresponding to discretization methods EF and EB, evaluation measure EMD, WRMSE and PE ATE, and inferring CPT method MLE or EBP. The color indicates the number of bins in the best approach for the corresponding experiment with respect to the chosen evaluation measure: light blue stands for a small number of bins and dark blue means a large number of bins. In the heatmap, a star indicates MDLP dominance over all solutions for that binning strategy, a plus signifies CAIM dominance, a minus denotes ChiMerge dominance and a tilde denotes dynamic discretization dominance.

Error measure		EMD	PE ATE	
CPT method		All	EBP	
Binning method		–	EF	EW
Experiment	N			
CQ DGP	2500	EF30	EF30	EW25
Lalonde	2675	EF12	EF9*	EW3*
MC	4000	EW30		

Table B.2: Heatmap summarizing all the non-causal results.

Error measure		EMD	WRMSE			
CPT method		NA	EBP	MLE	EBP	MLE
Binning method		All	EF	EF	EW	EW
Experiment	N					
LG9	5000	EF30	EF17	EF17	EW12	EW12*
LG10	3000	EF30	EF17*	EF17*	EW10*	EW10*
LG11	2000	EF30	EF12*	EF12*	EW8*	EW8*
LG12	1000	EF30	EF12	EF12	EW8*	EW8*
LG13	800	EF30	EF12	EF12	EW5*	EW5*
LG14	600	EW30	EF8	EF8	EW8*	EW8*
LG15	500	EW30	EF8	EF8	EW8*	EW8*
LG16	400	EW30	EF8	EF8	EW5*	EW5*
LG17	300	EW30	EF8	EF8	EW5*	EW5*
LG18	200	EW30	EF5	EF5	EW5-	EW5-
LG19	100	EW30	EF5	EF5	EW5-	EW5-
LG101	100	EW30	EF5	EF5	EW8*	EW8*
LG102	1050	EF30	EF12	EF12	EW8	EW8
LG103	1525	EF30	EF12	EF12	EW10*	EW10*
LG104	575	EF30	EF5*	EF5*	EW5*	EW5*
LG105	812	EF30	EF12	EF12	EW8*	EW8*
LG106	1762	EF30	EF17	EF17	EW8*	EW8*
LG107	1288	EF30	EF12	EF12	EW8*	EW8*
LG108	338	EW30	EF8	EF8	EW5-	EW5-
LG109	456	EW30	EF8	EF8	EW5*	EW5*
LG110	1406	EF30	EF10	EF10	EW8-	EW8-
LG111	1881	EF30	EF10*	EF10*	EW12*	EW12*
LG112	931	EF30	EF10	EF10	EW5*	EW5*
LG113	694	EW30	EF10	EF10	EW8*	EW8*
LG114	1644	EW30	EF14	EF14	EW8-	EW8-
LG115	1169	EF30	EF12	EF12	EW10*	EW10*
LG116	219	EW30	EF5	EF5	EW5-	EW5-
LG117	278	EW30	EF5	EF5	EW5-	EW5-
LG118	1228	EF30	EF10	EF10	EW8*	EW8*
LG119	1703	EF30	EF17	EF17	EW10*	EW10*
LG120	753	EF30	EF12	EF12	EW8-	EW8-
LG121	991	EF30	EF12*	EF12*	EW8*	EW8*
LG122	1941	EF30	EF10	EF10	EW8*	EW8*
LG123	1466	EF30	EF12	EF12	EW8*	EW8*
LG124	516	EW30	EF10	EF10	EW5*	EW5*
LG125	397	EW30	EF8	EF8	EW5*	EW5*
NM1	500	EW30+	EF25+	EF25+	EW20+	EW20+
NM2	500	EF30+	EF30+	EF30+	EW30+	EW30+
NM3	500	EF30+	EF25+	EF25+	EW17+	EW17+
NM4	500	EF30+	EF30+	EF30+	EW25+	EW25+
NM5	100	EF30+	EF8+	EF8+	EW17+	EW17+
NM6	100	EF30+	EF30+	EF30+	EW25+	EW25+
NM7	100	EF30+	EF17+	EF17+	EW30+	EW30+
NM8	100	EF25+	EF25+	EF25+	EW30+	EW30+
Arth	1000	EW30				

B.2 Experimental Set-up

The experimental setup outlines the specifications of each of the experiments in terms of their distribution and sample size.

Table B.3: The first 25 experiments involving linear Gaussian BNs parameterized by Sobol sequences for the number of samples N and standard deviations. Each variable follows a normal distribution $\mathcal{N}(\mu, \sigma)$, with mean μ specified in the header and standard deviation σ listed in the table.

		$P(X_1)$	$P(X_2)$	$P(X_3)$	$P(X_4 \mid X_1, X_2)$	$P(X_5 \mid X_3, X_4)$
	μ	20	20	15	$2X_1 + 3X_2$	$3X_3 + 3X_4$
Experiment	N	σ	σ	σ	σ	σ
LG101	100	1	1	1	1	1
LG102	1050	5.5	5.5	5.5	5.5	5.5
LG103	1525	3.25	3.25	3.25	3.25	3.25
LG104	575	7.75	7.75	7.75	7.75	7.75
LG105	813	4.38	4.38	4.38	4.38	4.38
LG106	1763	8.88	8.88	8.88	8.88	8.88
LG107	1288	2.13	2.13	2.13	2.13	2.13
LG108	338	6.63	6.63	6.63	6.63	6.63
LG109	456	3.81	3.81	3.81	3.81	3.81
LG110	1406	8.31	8.31	8.31	8.31	8.31
LG111	1881	1.56	1.56	1.56	1.56	1.56
LG112	931	6.06	6.06	6.06	6.06	6.06
LG113	694	2.69	2.69	2.69	2.69	2.69
LG114	1644	7.19	7.19	7.19	7.19	7.19
LG115	1169	4.94	4.94	4.94	4.94	4.94
LG116	219	9.44	9.44	9.44	9.44	9.44
LG117	278	5.22	5.22	5.22	5.22	5.22
LG118	1228	9.72	9.72	9.72	9.72	9.72
LG119	1703	2.97	2.97	2.97	2.97	2.97
LG120	754	7.47	7.47	7.47	7.47	7.47
LG121	991	1.84	1.84	1.84	1.84	1.84
LG122	1941	6.34	6.34	6.34	6.34	6.34
LG123	1466	4.09	4.09	4.09	4.09	4.09
LG124	516	8.59	8.59	8.59	8.59	8.59
LG125	397	2.41	2.41	2.41	2.41	2.41

Table B.4: Number of samples and parametrization of the experiments with the 11 extra linear Gaussian Bayesian networks. These experiments are meant to isolate the effect of the sample size on the Pareto front. Note that the samples in lower sample-sized experiments are contained in the samples of experiments with higher sample sizes.

(a) Parametrization of all of the experiments

$P(X_1)$	$P(X_2)$	$P(X_3)$	$P(X_4 \mid X_1, X_2)$	$P(X_5 \mid X_3, X_4)$
$\mathcal{N}(20, 2)$	$\mathcal{N}(20, 2)$	$\mathcal{N}(15, 2)$	$\mathcal{N}(2X_1 + 3X_2, 2)$	$\mathcal{N}(3X_3 + 3X_4, 2)$

(b) Name and sample size of experiments

Experiment	**LG9**	**LG10**	**LG11**	**LG12**	**LG13**	**LG14**	**LG15**	**LG16**	**LG17**	**LG18**	**LG19**
N	5000	3000	2000	1000	800	600	500	400	300	200	100

Table B.5: Name, number of samples, and parametrization of the experiments with the Normal mixture model

Experiment	N	$P(X_1)$	$P(X_2 \mid X_1 = 1)$	$P(X_2 \mid X_1 = 0)$
NM1	500	$B(1, \frac{1}{2})$	$\mathcal{N}(21, 10)$	$\mathcal{N}(25, 1)$
NM2	500	$B(1, \frac{4}{5})$	$\mathcal{N}(21, 10)$	$\mathcal{N}(25, 1)$
NM3	500	$B(1, \frac{1}{2})$	$\mathcal{N}(6, 2)$	$\mathcal{N}(4, 2)$
NM4	500	$B(1, \frac{4}{5})$	$\mathcal{N}(6, 2)$	$\mathcal{N}(4, 2)$
NM5	100	$B(1, \frac{1}{2})$	$\mathcal{N}(21, 10)$	$\mathcal{N}(25, 1)$
NM6	100	$B(1, \frac{4}{5})$	$\mathcal{N}(21, 10)$	$\mathcal{N}(25, 1)$
NM7	100	$B(1, \frac{1}{2})$	$\mathcal{N}(6, 2)$	$\mathcal{N}(4, 2)$
NM8	100	$B(1, \frac{4}{5})$	$\mathcal{N}(6, 2)$	$\mathcal{N}(4, 2)$

The following causal quadratic (CQ) experiment is adopted from Parikh et al. [176]:

$$
\begin{aligned}
&X_i \sim \mathcal{N}(0, 1) \\
&Y_i(0) = \beta^T X_i + \epsilon_0 && \text{where } \epsilon_0 \sim \mathcal{N}(0, 1) \\
&Y_i(1) = Y_i(0)^2 + \alpha^T X_i + \epsilon_1 && \text{where } \epsilon_1 \sim \mathcal{N}(0, 1) \\
&T_i = \text{expit}(\mathbf{1}^T X_i))
\end{aligned}
$$

A maximum number of 30 bins was allowed for all experiments except for the Lalonde dataset (12 bins maximum).

B.3 Evaluation Measures

The conditional expected value of the target variable, $\mathbb{E}[Y \mid X]$, in the original Bayesian network is compared with its counterpart in the discretised network, denoted $\mathbb{E}_{\text{disc}}[Y \mid X]$, where X is a root node. To account for the distribution of X, the accuracy of the discretized expectation is assessed using the weighted root mean squared error (WRMSE):

$$WRMSE = \sqrt{\sum_{x} P(X = x)(\mathbb{E}[Y \mid X = x] - \mathbb{E}_{disc}[Y \mid X = x])^2}$$

where $P(X = x)$ denotes the discretized probability that X takes value x. Note that the number of values involved in WRMSE depends on the discretization of the root node X.

For the causal Bayesian networks in Section 5.3.2, the average treatment effect is obtained using the adjustment formula detailed in Chapter 3. The experiments investigate the percentage error (PE) of the average treatment effect (ATE) as the primary object of analysis:

$$PE = 100 \times \left| \frac{ATE_{true} + ATE_{disc}}{ATE_{true}} \right|.$$

Appendix C

Appendix Chapter 6

C.1 Experimental Data of Deterring Hybrid Threat

This section specifies the details about the experiments of the experimental section by specifying the probability distributions these experiments are drawn from.

The most damaging impact θ_1 and the substantial impact θ_2 are drawn from the truncated normal distributions $f(x, \mu_1, \sigma_1, a, b)$ and $f(x, \mu_2, \sigma_2, a, b)$ respectively, where $a = 0$, $\mu_1, \mu_2 = 1000, 100$ and $\sigma_1, \sigma_2 = 300, 50$ respectively [170]. Hence x is drawn from the interval $[0, \infty]$. Negligible damaging impact θ_3 is drawn from a positive half-normal distribution based on normal distribution $\mathcal{N}(0, 5)$. All values represent millions of damaging costs.

The costs γ_i for each of the counter-hybrid measures d_i are drawn from truncated normal distributions. The probability of the adversary conducting a hybrid operation q_{ij} after counter-hybrid measure d_i is drawn from beta distributions as it is the conjugate prior to the Bernoulli distribution (assuming the adversary either attacks or does not attack). Finally, the probabilities of the potential impact of hybrid conduct w_{ijh} based on counter-hybrid measure d_i and attack c_j are drawn from a Dirichlet distribution as it is the conjugate prior to the categorical distribution and commonly used in influence diagrams [228].

Active intelligence sharing. When intelligence is shared, this is not directly communicated to the adversary, leaving the probability that the adversary conducts a hybrid operation after this measure q_{11} at $Be(5, 5)$. However, this measure significantly improves the mitigation ability of the damaging impact of the hybrid conduct

leaving the probabilities being drawn from Dirichlet distribution:

$$\begin{pmatrix} w_{111} \\ w_{112} \\ w_{113} \end{pmatrix} \sim \text{Dir} \begin{pmatrix} 4 \\ 8 \\ 12 \end{pmatrix}.$$

The cost of this measure accounts for losing confidential information to unauthorized parties and is drawn from the truncated normal distributions $f(x, \mu, \sigma, a, b)$ with $\mu = 150, \sigma = 50$ and $a = 0$. In case the adversary does not conduct a hybrid operation, the probability that the damaging impact will be negligible is always 1.

Boost cyber resilience at the wider societal level. Boosting cyber resilience works via deterrence by denial, leaving the probability that the measure dissuades the adversary from committing a hybrid conduct q_{21} drawn from $Be(4, 8)$ and the mitigation of potential impact drawn from the following Dirichlet distributions:

$$\begin{pmatrix} w_{211} \\ w_{212} \\ w_{213} \end{pmatrix} \sim \text{Dir} \begin{pmatrix} 3 \\ 5 \\ 8 \end{pmatrix}.$$

The cost of this measure is mostly carried by the private sector and is drawn from truncated normal distribution $f(x, \mu, \sigma, a, b)$ with $\mu = 300, \sigma = 50$ and $a = 0$.

Offensive cyber operation. An offensive cyber operation has deterrence by denial as well as deterrence by punishment components. To compensate for the fact that the measure can also backfire, the probability that this measure is successful in dissuading the adversary from committing hybrid conduct q_{31} is drawn from $Be(1, 1.2)$ and the damage mitigation potential drawn from the following Dirichlet distributions:

$$\begin{pmatrix} w_{311} \\ w_{312} \\ w_{313} \end{pmatrix} \sim \text{Dir} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

The costs of this measure contain the cost of setting up an offensive cyber unit as well as the cost that comes with the associated attack. It is drawn from truncated normal distribution $f(x, \mu, \sigma, a, b)$ with $\mu = 250, \sigma = 30$ and $a = 0$.

Market restriction. Imposing market restrictions works via deterrence by denial and therefore the probability that this counter-hybrid measure dissuades the adversary from committing hybrid conduct q_{41} is drawn from $Be(2, 8)$. The probabilities of potential impacts are drawn from the Dirichlet distribution

$$\begin{pmatrix} w_{411} \\ w_{412} \\ w_{413} \end{pmatrix} \sim \text{Dir} \begin{pmatrix} 2 \\ 2 \\ 15 \end{pmatrix}.$$

Finally, the costs of the measure involve excluding certain private organizations from the market and is drawn from truncated normal distribution $f(x, \mu, \sigma, a, b)$ with $\mu = 400, \sigma = 50$ and $a = 0$.

Open deterrence messaging through strategic communications. Assuming that the message involves some deterrence by punishment and there is uncertainty involved in threatening, the probability that this deterrence measure successfully dissuades the adversary from committing a hybrid conduct q_{51} is drawn from $Be(0.4, 2)$. Damage mitigation is not involved and therefore the probabilities for damaging impacts are drawn from the same Dirichlet distribution as no measure was taken:

$$\begin{pmatrix} w_{511} \\ w_{512} \\ w_{513} \end{pmatrix} \sim \text{Dir} \begin{pmatrix} 12 \\ 6 \\ 2 \end{pmatrix}.$$

Finally, the costs of the measure also include costs of a risky escalation that one commits to and is drawn from truncated normal distribution $f(x, \mu, \sigma, a, b)$ with $\mu = 500, \sigma = 250$ and $a = 0$.

No deterrence measure. Assuming no deterrence measure taken d_6, the probability that the adversary conducts a hybrid operation q_{61} is drawn from $Be(5, 5)$. Similarly, in case the defender does not conduct a deterrence effort and the adversary conducts a hybrid operation the probability of each of the three impacts is drawn from the Dirichlet distribution:

$$\begin{pmatrix} w_{611} \\ w_{612} \\ w_{613} \end{pmatrix} \sim \text{Dir} \begin{pmatrix} 12 \\ 6 \\ 2 \end{pmatrix}.$$

In case the adversary does not conduct a hybrid operation, the probability that the damaging impact will be negligible is always 1.

Figure C.1 and C.2 illustrate the distribution of successful deterrence and the distribution of the potential impact of the malicious cyber attack, respectively.

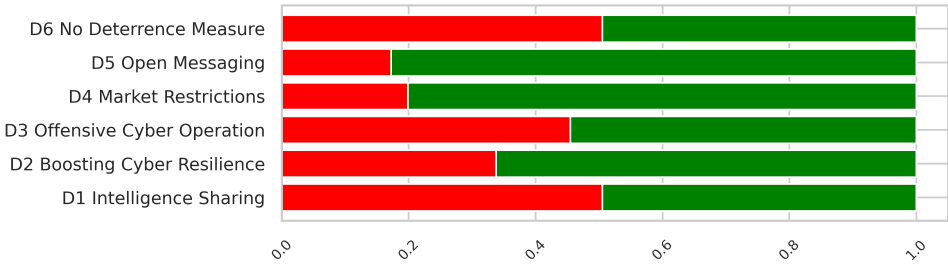


Figure C.1: The probability that each of the counter-hybrid measures succeeds in deterring the adversary sampled from Beta Distribution. While the green indicates the probability that the adversary is successfully dissuaded, the red illustrates the probability that the adversary still conducts a cyber operation.

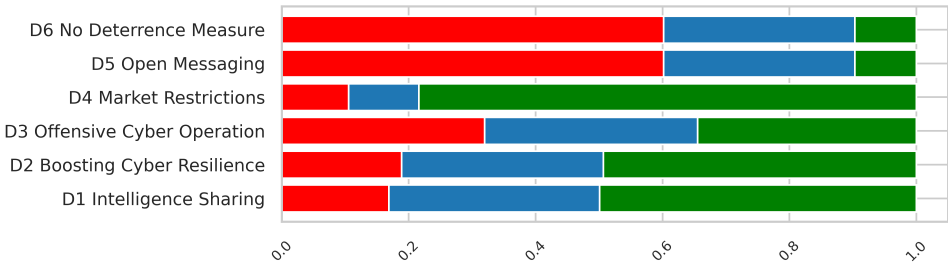


Figure C.2: The probability that each of the counter-hybrid measures succeeds in mitigating the hybrid operation sampled from Dirichlet distribution. While the red indicates the probability the hybrid operation has a severe impact, the blue indicates the operation has a mediocre impact and the green indicates the hybrid operation has negligible impact.

Bibliography

[1] John T Abatzoglou, Solomon Z Dobrowski, Sean A Parks, and Katherine C Hegewisch. Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific data*, 5(1):1–12, 2018.

[2] Daron Acemoglu, Leopoldo Fergusson, and Simon Johnson. Population and conflict. *The Review of Economic Studies*, 87(4):1565–1604, 2020.

[3] Wario R Adano, Ton Dietz, Karen Witsenburg, and Fred Zaal. Climate change, violent conflict and local institutions in kenya's drylands. *Journal of peace research*, 49(1):65–80, 2012.

[4] Virginia Aglietti, Xiaoyu Lu, Andrei Paleyes, and Javier González. Causal bayesian optimization. In *AISTATS*, pages 3155–3164. PMLR, 2020.

[5] Ibrahim Alkhairy, Samantha Low-Choy, Justine Murray, Junhu Wang, and Anthony Pettitt. Quantifying conditional probability tables in bayesian networks: Bayesian regression for scenario-based encoding of elicited expert assessments on feral pig habitat. *Journal of Applied Statistics*, 47(10):1848–1884, 2020.

[6] Pierre Allan and Cédric Dupont. International relations theory and game theory: Baroque modeling choices and empirical robustness. *International Political Science Review*, 20(1):23–47, 1999.

[7] Holly Andersen. When to expect violations of causal faithfulness and why it matters. *Philosophy of Science*, 80(5):672–683, 2013.

[8] Steen A Andersson, David Madigan, and Michael D Perlman. Alternative markov properties for chain graphs. *Scandinavian Journal of Statistics*, 28(1):33–85, 2001.

[9] Alex Andrew, Sam Spillard, Joshua Collyer, and Neil Dhir. Developing optimal causal cyber-defence agents via cyber security simulation. *arXiv preprint arXiv:2207.12355*, 2022.

[10] Bryan Andrews, Peter Spirtes, and Gregory F Cooper. On the completeness of causal discovery in the presence of latent confounding with tiered background knowledge. In *International Conference on Artificial Intelligence and Statistics*, pages 4002–4011. PMLR, 2020.

[11] Ankur Ankan and Abinash Panda. pgmpy: Probabilistic graphical models using python. In *Proceedings of the 14th Python in Science Conference (SCIPY 2015)*. Citeseer, 2015.

[12] David Applegate, Tamraparni Dasu, Shankar Krishnan, and Simon Urbanek. Unsupervised clustering of multidimensional distributions using earth mover distance. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 636–644, 2011.

[13] David Arbour, Dan Garant, and David Jensen. Inferring network effects from observational data. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 715–724, New York, NY, USA, 2016. Association for Computing Machinery.

[14] Daniel Arce. Cybersecurity for defense economists. *Defence and Peace Economics*, 34(6):705–725, 2023.

[15] Peter M. Aronow and Cyrus Samii. Estimating average causal effects under general interference, with application to a social network experiment. *Annals of Applied Statistics*, 11(4):1912–1947, December 2017.

[16] Konstantin Ash and Nick Obradovich. Climatic stress, internal migration, and syrian civil war onset. *Journal of Conflict Resolution*, 64(1):3–31, 2020.

[17] Charles K Assaad, Emilie Devijver, and Eric Gaussier. Survey and evaluation of causal discovery methods for time series. *Journal of Artificial Intelligence Research*, 73:767–819, 2022.

[18] Afraa Attiah, Mainak Chatterjee, and Cliff C Zou. A game theoretic approach to model cyber attack and defense strategies. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–7. IEEE, 2018.

[19] Thomas Bäck and et al. Evolutionary algorithms for parameter optimization—thirty years later. *Evolutionary Computation*, 31(2):81–122, 2023.

[20] Günther Baechler. Conclusions: future relevance and priorities of small states. In *Small States Inside and Outside the European Union: Interests and Policies*, pages 267–283. Springer, 1998.

[21] Ying Bai and James Kai-sing Kung. Climate shocks and sino-nomadic conflict. *Review of Economics and Statistics*, 93(3):970–981, 2011.

[22] Elnaz Bajoori, János Flesch, and Dries Vermeulen. Behavioral perfect equilibrium in bayesian games. *Games and Economic Behavior*, 98:78–109, 2016.

[23] Michelle Baker and Terrance E Boult. Pruning bayesian networks for efficient computation. *arXiv preprint arXiv:1304.1112*, 2013.

[24] Mariusz Balaban and Paweł Mielniczek. Hybrid conflict modeling. In *2018 Winter Simulation Conference (WSC)*, pages 3709–3720. IEEE, 2018.

[25] Pieter Balcaen, Cind Du Bois, and Caroline Buts. A game-theoretic analysis of hybrid threats. *Defence and Peace Economics*, 33(1):26–41, 2022.

[26] Elias Bareinboim, Carlos Brito, and Judea Pearl. Local characterizations of causal bayesian networks. In *Graph Structures for Knowledge Representation and Reasoning*, pages 1–17, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[27] Elias Bareinboim, Juan David Correa, Duligur Ibeling, and Thomas F. Icard. On pearl's hierarchy and the foundations of causal inference. *Probabilistic and Causal Inference*, 2022.

[28] Elias Bareinboim and Judea Pearl. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences*, 113(27):7345–7352, 2016.

[29] Martine J Barons, Steven Mascaro, and Anca M Hanea. Balancing the elicitation burden and the richness of expert input when quantifying discrete bayesian networks. *Risk Analysis*, 42(6):1196–1234, 2022.

[30] Jeffrey D Berejikian. A cognitive theory of deterrence. *Journal of Peace Research*, 39(2):165–183, 2002.

[31] Tomas Beuzen, Lucy Marshall, and Kristen D Splinter. A comparison of methods for discretizing continuous variables in bayesian networks. *Environmental modelling & software*, 108:61–66, 2018.

[32] Hans-Georg Beyer and Hans-Paul Schwefel. Evolution strategies–a comprehensive introduction. *Natural computing*, 1:3–52, 2002.

[33] Rohit Bhattacharya, Daniel Malinsky, and Ilya Shpitser. Causal inference under interference and network uncertainty. In *Uncertainty in Artificial Intelligence*, pages 1028–1038. PMLR, 2020.

[34] Rohit Bhattacharya, Razieh Nabi, and Ilya Shpitser. Semiparametric inference for causal effects in graphical models with hidden variables. *Journal of Machine Learning Research*, 23(295):1–76, 2022.

[35] Rohit Bhattacharya, Tushar Nagarajan, Daniel Malinsky, and Ilya Shpitser. Differentiable causal discovery under unmeasured confounding. In *International Conference on Artificial Intelligence and Statistics*, pages 2314–2322. PMLR, 2021.

[36] Stephan Bongers, Patrick Forré, Jonas Peters, and Joris M Mooij. Foundations of structural causal models with cycles and latent variables. *The Annals of Statistics*, 49(5):2885–2915, 2021.

[37] Craig Boutilier, Nir Friedman, Moises Goldszmidt, and Daphne Koller. Context-specific independence in bayesian networks. In *Proceedings of the Twelfth International Conference on Uncertainty in Artificial Intelligence*, UAI'96, page 115–123, San Francisco, CA, USA, 1996. Morgan Kaufmann Publishers Inc.

[38] Aaron F. Brantly. The cyber deterrence problem. In *2018 10th International Conference on Cyber Conflict (CyCon)*, pages 31–54, 2018.

[39] Charles Broyden. The Convergence of a Class of Double-rank Minimization Algorithms: 2. The New Algorithm. *IMA Journal of Applied Mathematics*, 6(3), 09 1970.

[40] Randal E Bryant. Graph-based algorithms for Boolean function manipulation. *Transactions on Computers*, 100(8):677–691, 1986.

[41] Simone Busetti. Causality is good for practice: policy design and reverse engineering. *Policy Sciences*, 56(2):419–438, 2023.

[42] Federico Carli, Manuele Leonelli, Eva Riccomagno, and Gherardo Varando. The r package stagedtrees for structural learning of stratified staged trees. *Journal of Statistical Software*, 102:1–30, 2022.

[43] Nancy Cartwright. Causal diversity and the markov condition. *Synthese*, 121(1/2):3–27, 1999.

[44] Lars-Erik Cederman and Nils B Weidmann. Predicting armed conflict: Time to adjust our expectations? *Science*, 355(6324):474–476, 2017.

[45] R. L. Chambers, Hugh Kennedy, John E. Woods, Majid Khadduri, and Gerald Henry Blake. Iraq. *Encyclopedia Britannica*, 2024.

[46] Mark Chavira and Adnan Darwiche. On probabilistic inference by weighted model counting. *Artificial Intelligence*, 172(6-7):772–799, 2008.

[47] Debo Cheng, Jiuyong Li, Lin Liu, Jixue Liu, and Thuc Duy Le. Data-driven causal effect estimation based on graphical causal modelling: A survey. *ACM Computing Surveys*, 56(5):1–37, 2024.

[48] David Maxwell Chickering. Optimal structure identification with greedy search. *Journal of Machine Learning Research*, 3:507–554, mar 2003.

[49] John Y. Ching, Andrew K. C. Wong, and Keith C. C. Chan. Class-dependent discretization for inductive learning from continuous and mixed-mode data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7):641–651, 1995.

[50] SEDAC CIESIN. Gridded population of the world, version 4 (gpwv4): population density. *Center for International Earth Science Information Network-CIESIN-Columbia University. NASA Socioeconomic Data and Applications Center (SEDAC). https://doi. org/10.7927/H4DZ068D*, 2015.

[51] Stephen R Cole and Constantine E Frangakis. The consistency statement in causal inference: a definition or an assumption? *Epidemiology*, 20(1):3–5, 2009.

[52] Diego Colombo, Marloes H. Maathuis, Markus Kalisch, and Thomas S. Richardson. Learning high-dimensional directed acyclic graphs with latent and selection variables. *The Annals of Statistics*, pages 294–321, 2012.

[53] Juan Correa and Elias Bareinboim. A calculus for stochastic interventions:causal effect identification and surrogate experiments. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(06):10093–10100, Apr. 2020.

[54] David Roxbee Cox. *Planning of experiments.* Wiley, 1958.

[55] Peter M Cox, Richard A Betts, Chris D Jones, Steven A Spall, and Ian J Totterdell. Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, 408(6809):184–187, 2000.

[56] Giso H. Dal, Alfons W. Laarman, Arjen Hommersom, and Peter J.F. Lucas. A compositional approach to probabilistic knowledge compilation. *International Journal of Approximate Reasoning*, 138:38–66, 2021.

[57] Giso H Dal and Peter JF Lucas. Weighted positive binary decision diagrams for exact probabilistic inference. *International Journal of Approximate Reasoning*, 90:411–432, 2017.

[58] David Danks. *Unifying the mind: Cognitive representations as graphical models.* Mit Press, 2014.

[59] Adnan Darwiche and Pierre Marquis. A knowledge compilation map. *Journal of Artificial Intelligence Research*, 17:229–264, 2002.

[60] Paul K Davis, Angela O'Mahony, Christian Curriden, and Jonathan Lamb. Influencing adversary states: Quelling perfect storm. Technical report, RAND CORP SANTA MONICA CA, 2021.

[61] A Philip Dawid. Causal inference without counterfactuals. *Journal of the American Statistical Association*, 95(450):407–424, 2000.

[62] A. Philip Dawid. Beware of the dag! In *Proceedings of Workshop on Causality: Objectives and Assessment at NIPS 2008*, volume 6, pages 59–86, Whistler, Canada, 12 Dec 2010. PMLR.

[63] Jacob de Nobel, Furong Ye, Diederick Vermetten, Hao Wang, Carola Doerr, and Thomas Bäck. IOHexperimenter: Benchmarking platform for iterative optimization heuristics. *CoRR*, abs/2111.04077, 2021.

[64] Rajeev H Dehejia and Sadek Wahba. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American statistical Association*, 94(448):1053–1062, 1999.

[65] David-Joaquín Delgado-Hernández, Oswaldo Morales-Nápoles, David De-León-Escobedo, and Juan-Carlos Arteaga-Arcos. A continuous bayesian network for earth dams' risk assessment: an application. *Structure and Infrastructure Engineering*, 10(2):225–238, 2014.

[66] Tom Deligiannis. The evolution of environment-conflict research: Toward a livelihood framework. *Global Environmental Politics*, 12(1):78–100, 2012.

[67] Eliana Duarte and Liam Solus. Representation of context-specific causal models with observational and interventional data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, page qkaf059, 2025.

[68] Alexander D'Amour, Peng Ding, Avi Feller, Lihua Lei, and Jasjeet Sekhon. Overlap in observational studies with high-dimensional covariates. *Journal of Econometrics*, 221(2):644–654, 2021.

[69] Frederick Eberhardt. Introduction to the epistemology of causation. *Philosophy Compass*, 4(6):913–925, 2009.

[70] Frederick Eberhardt and Richard Scheines. Interventions and causal inference. *Philosophy of Science*, 74(5):981–995, 2007.

[71] Usama M. Fayyad and Keki B. Irani. Multi-interval discretization of continuous-valued attributes for classification learning. In *International Joint Conference on Artificial Intelligence*, 1993.

[72] Hanne Fjelde and Nina Von Uexkull. Climate triggers: Rainfall anomalies, vulnerability and communal conflict in sub-saharan africa. *Political Geography*, 31(7):444–453, 2012.

[73] Roger Fletcher. A new approach to variable metric algorithms. *The Computer Journal*, 13(3):317–322, 01 1970.

[74] Malcolm Forster, Garvesh Raskutti, Reuben Stern, and Naftali Weinberger. The frugal inference of causal relations. *The British Journal for the Philosophy of Science*, 69(3):821–848, 2018.

[75] James Fox, Tom Everitt, Ryan Carey, Eric D Langlois, Alessandro Abate, and Michael J Wooldridge. Pycid: A python library for causal influence diagrams. In *SciPy*, pages 65–73, 2021.

[76] Drew Fudenberg and Jean Tirole. Game theory mit press. *Cambridge, MA*, 86, 1991.

[77] Salvador García, Julián Luengo, José Antonio Sáez, Victoria López, and Francisco Herrera. A survey of discretization techniques: Taxonomy and empirical analysis in supervised learning. *IEEE Transactions on Knowledge and Data Engineering*, 25(4):734–750, 2013.

[78] Sushant S Garud, Iftekhar A Karimi, and Markus Kraft. Design of computer experiments: A review. *Computers & Chemical Engineering*, 106:71–95, 2017.

[79] Tomas Geffner and et al. Deep end-to-end causal inference. *arXiv preprint arXiv:2202.02195*, 2022.

[80] Dan Geiger and David Heckerman. Knowledge representation and inference in similarity networks and bayesian multinets. *Artificial Intelligence*, 82(1):45–74, 1996.

[81] Ramesh Ghimire and Susana Ferreira. Floods and armed conflict. *Environment and Development Economics*, 21(1):23–52, 2016.

[82] Amitai Gilad and Asher Tishler. Mitigating the risk of advanced cyber attacks: The role of quality, covertness and intensity of use of cyber weapons. *Defence and peace economics*, ahead-of-print(ahead-of-print):1–21, 2023.

[83] Herbert Gintis. Beyond homo economicus: evidence from experimental economics. *Ecological Economics*, 35(3):311–322, 2000.

[84] Donald Goldfarb. A family of variable-metric methods derived by variational means. *Mathematics of Computation*, 24(109):23–26, 1970.

[85] Irving John Good. Some history of the hierarchical bayesian methodology. *Trabajos de estadística y de investigación operativa*, 31:489–519, 1980.

[86] Olivier Goudet, Diviyan Kalainathan, Michèle Sebag, and Isabelle Guyon. Learning bivariate functional causal models. In *Cause Effect Pairs in Machine Learning*, pages 101–153. Springer, 2019.

[87] Paula Gradu, Tijana Zrnic, Yixin Wang, and Michael I Jordan. Valid inference after causal discovery. *Journal of the American Statistical Association*, 0(0):1–21, 2024.

[88] Andy Greenberg. *Sandworm: A new era of cyberwar and the hunt for the Kremlin's most dangerous hackers*. Anchor, 2019.

[89] Ruocheng Guo, Jundong Li, and Huan Liu. Learning individual causal effects from networked observational data. In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 232–240, 2020.

[90] Lewis Hammond, James Fox, Tom Everitt, Alessandro Abate, and Michael Wooldridge. Equilibrium refinements for multi-agent influence diagrams: theory and practice. *arXiv preprint arXiv:2102.05008*, 2021.

[91] Lewis Hammond, James Fox, Tom Everitt, Ryan Carey, Alessandro Abate, and Michael Wooldridge. Reasoning about causality in games. *Artificial Intelligence*, 320:103919, 2023.

[92] John C Harsanyi. Games with incomplete information played by "bayesian" players, i–iii part i. the basic model. *Management science*, 14(3):159–182, 1967.

[93] Jason Hartford, Greg Lewis, Kevin Leyton-Brown, and Matt Taddy. Deep iv: A flexible approach for counterfactual prediction. In *International Conference on Machine Learning*, Proceedings of Machine Learning Research, pages 1414–1423. PMLR, 2017.

[94] Uzma Hasan and Md Osman Gani. Kcrl: A prior knowledge based causal discovery framework with reinforcement learning. In Zachary Lipton, Rajesh Ranganath, Mark Sendak, Michael Sjoding, and Serena Yeung, editors, *Proceedings of the 7th Machine Learning for Healthcare Conference*, volume 182 of *Proceedings of Machine Learning Research*, pages 691–714. PMLR, 05–06 Aug 2022.

[95] Kirsty L Hassall, Gordon Dailey, Joanna Zawadzka, Alice E Milne, Jim A Harris, Ron Corstanje, and Andrew P Whitmore. Facilitating the elicitation of beliefs for use in bayesian belief modelling. *Environmental Modelling & Software*, 122:104539, 2019.

[96] Alain Hauser and Peter Bühlmann. Jointly interventional and observational data: estimation of interventional markov equivalence classes of directed acyclic graphs. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 77(1):291–318, 2015.

[97] Daniel M Hausman and James Woodward. Independence, invariance and the causal markov condition. *The British Journal for the Philosophy of Science*, 50(4):521–583, 1999.

[98] Cullen S Hendrix and Sarah M Glaser. Trends and triggers: Climate, climate change and civil conflict in sub-saharan africa. *Political geography*, 26(6):695–715, 2007.

[99] Leon Hermans, Scott Cunningham, and Jill Slinger. The usefulness of game theory as a method for policy evaluation. *Evaluation*, 20(1):10–25, 2014.

[100] M.A. Hernan and J.M. Robins. *Causal Inference: What If*. Chapman & Hall/CRC Monographs on Statistics & Applied Probab. CRC Press, 2020.

[101] Edwin Ho, Arvind Rajagopalan, Alex Skvortsov, Sanjeev Arulampalam, and Mahendra Piraveenan. Game theory in defence applications: a review. *Sensors*, 22(3):1032, 2022.

[102] Paul W Holland. Statistics and causal inference. *Journal of the American Statistical Association*, 81(396):945–960, 1986.

[103] Thomas F Homer-Dixon. On the threshold: environmental changes as causes of acute conflict. *International security*, 16(2):76–116, 1991.

[104] Thomas F Homer-Dixon. Environmental scarcities and violent conflict: evidence from cases. *International security*, 19(1):5–40, 1994.

[105] Thomas F Homer-Dixon. *Environment, scarcity, and violence*. Princeton University Press, 2010.

[106] Thomas F Homer-Dixon and Jessica Blitt. *Ecoviolence: Links among environment, population and security*. Rowman & Littlefield, 1998.

[107] Nigel Howard. Paradoxes of rationality. *Cambridge, Mass*, page 48f, 1971.

[108] Ronald A Howard and James E Matheson. Influence diagrams. *Decision Analysis*, 2(3):127–143, 2005.

[109] Patrik Hoyer, Dominik Janzing, Joris M Mooij, Jonas Peters, and Bernhard Schölkopf. Nonlinear causal discovery with additive noise models. *Advances in Neural Information Processing Systems*, 21, 2008.

[110] Patrik O Hoyer, Shohei Shimizu, Antti J Kerminen, and Markus Palviainen. Estimation of causal effects using linear non-gaussian causal models with hidden variables. *International Journal of Approximate Reasoning*, 49(2):362–378, 2008.

[111] Solomon M Hsiang, Marshall Burke, and Edward Miguel. Quantifying the influence of climate on human conflict. *Science*, 341(6151):1235367, 2013.

[112] Solomon M Hsiang, Kyle C Meng, and Mark A Cane. Civil conflicts are associated with the global climate. *Nature*, 476(7361):438–441, 2011.

[113] Michael G Hudgens and M Elizabeth Halloran. Toward causal inference with interference. *Journal of the American Statistical Association*, 103(482):832–842, 2008.

[114] George Huffman. Imerg v06 quality index. *NASA, Available for download from: https://gpm. nasa. gov/sites/default/files/2020-02/IMERGV06_QI_0. pdf*, 2019.

[115] Antti Hyttinen, Johan Pensar, Juha Kontinen, and Jukka Corander. Structure learning for bayesian networks over labeled dags. In *Proceedings of the Ninth International Conference on Probabilistic Graphical Models*, volume 72 of *Proceedings of Machine Learning Research*, pages 133–144. PMLR, 2018.

[116] Tobias Ide, Miguel Rodriguez Lopez, Christiane Fröhlich, and Jürgen Scheffran. Pathways to water conflict during drought in the mena region. *Journal of Peace Research*, 58(3):568–582, 2021.

[117] Guido W. Imbens and Joshua D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994.

[118] Guido W. Imbens and Donald B. Rubin. *Rubin Causal Model*, pages 229–241. Palgrave Macmillan UK, London, 2010.

[119] Guido W Imbens and Donald B Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge University Press, 2015.

[120] James Fox, Tom Everitt, Ryan Carey, Eric Langlois, Alessandro Abate, and Michael Wooldridge. PyCID: A Python Library for Causal Influence Diagrams. In Meghann Agarwal, Chris Calloway, Dillon Niederhut, and David Shupe, editors, *Proceedings of the 20th Python in Science Conference*, pages 43 – 51, 2021.

[121] Zhiwei Ji, Qibiao Xia, and Guanmin Meng. A review of parameter learning methods in bayesian network. In *Advanced Intelligent Computing Theories and Applications: 11th International Conference, ICIC 2015, Fuzhou, China, August 20-23, 2015. Proceedings, Part III 11*, pages 3–12. Springer, 2015.

[122] Menglin Jin and Robert E Dickinson. Land surface skin temperature climatology: Benefitting from the strengths of satellite observations. *Environmental research letters*, 5(4):044004, 2010.

[123] Fredrik Johansson and Goran Falkman. A bayesian network approach to threat evaluation with application to an air defense scenario. In *2008 11th International conference on information fusion*, pages 1–7. IEEE, 2008.

[124] Randy Kerber. Chimerge: Discretization of numeric attributes. In *Proceedings of the tenth national conference on Artificial intelligence*, pages 123–128, 1992.

[125] David Kilcullen. The evolution of unconventional warfare. *Scandinavian Journal of Military Studies*, 2(1), 2019.

[126] Neville K Kitson and Anthony C Constantinou. The impact of variable ordering on bayesian network structure learning. *Data Mining and Knowledge Discovery*, pages 1–25, 2024.

[127] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[128] Daphne Koller and Brian Milch. Multi-agent influence diagrams for representing and solving games. *Games and Economic Behavior*, 45(1):181–221, 2003. First World Congress of the Game Theory Society.

[129] Ore Koren. Food resources and strategic conflict. *Journal of Conflict Resolution*, 63(10):2236–2261, 2019.

[130] Nadiya Kostyuk. Deterrence in the cyber realm: Public versus private cyber capacity. *International Studies Quarterly*, 65(4):1151–1162, 2021.

[131] Vally Koubi, Thomas Bernauer, Anna Kalbhenn, and Gabriele Spilker. Climate variability, economic growth, and civil conflict. *Journal of peace research*, 49(1):113–127, 2012.

[132] Rainer Kress. *Numerical analysis*, volume 181. Springer Science & Business Media, 2012.

[133] L.A. Kurgan and K.J. Cios. Caim discretization algorithm. *IEEE transactions on knowledge and data engineering*, 16(2):145–153, 2004.

[134] Gustavo Lacerda, Peter L Spirtes, Joseph Ramsey, and Patrik O Hoyer. Discovering cyclic causal models by independent components analysis. *arXiv preprint arXiv:1206.3273*, 2012.

[135] Robert J LaLonde. Evaluating the econometric evaluations of training programs with experimental data. *The American economic review*, pages 604–620, 1986.

[136] Eric Lamb, Steven Shirtliffe, and William May. Structural equation modeling in the plant sciences: An example using yield components in oat. *Canadian Journal of Plant Science*, 91(4):603–619, 2011.

[137] Eric G Lamb, Kerrie L Mengersen, Katherine J Stewart, Udayanga Attanayake, and Steven D Siciliano. Spatially explicit structural equation modeling. *Ecology*, 95(9):2434–2442, 2014.

[138] Steffen L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.

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

[140] Jaron JR Lee, Rohit Bhattacharya, Razieh Nabi, and Ilya Shpitser. Ananke: A python package for causal inference using graphical models. *arXiv preprint arXiv:2301.11477*, 2023.

[141] Sanghack Lee and Elias Bareinboim. Structural causal bandits with non-manipulable variables. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):4164–4172, Jul. 2019.

[142] Manuele Leonelli and Gherardo Varando. Context-specific causal discovery for categorical data using staged trees. In *International conference on artificial intelligence and statistics*, pages 8871–8888. PMLR, 2023.

[143] Kevin W Li, Keith W Hipel, D Marc Kilgour, and Liping Fang. Preference uncertainty in the graph model for conflict resolution. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 34(4):507–520, 2004.

[144] R.H.A. Lindelauf, H.J.M. Hamers, and B.G.M. Husslage. Cooperative game theoretic centrality analysis of terrorist networks: The cases of jemaah islamiyah and al qaeda. *European Journal of Operational Research*, 229(1):230–238, 2013.

[145] Andrew M Linke and Brett Ruether. Weather, wheat, and war: Security implications of climate variability for conflict in syria. *Journal of Peace Research*, 58(1):114–131, 2021.

[146] Piotr Lis and Jacob Mendel. Cyberattacks on critical infrastructure: an economic perspective 1. *Economics and Business Review*, 5(2):24–47, 2019.

[147] Christos Louizos, Uri Shalit, Joris M Mooij, David Sontag, Richard Zemel, and Max Welling. Causal effect inference with deep latent-variable models. *Advances in Neural Information Processing Systems*, 30, 2017.

[148] Yu Luo, Daniel J Graham, and Emma J McCoy. Semiparametric bayesian doubly robust causal estimation. *Journal of Statistical Planning and Inference*, 225:171–187, 2023.

[149] Ashique Mahmood. Structure learning of causal bayesian networks: A survey, 2011.

[150] Marc Maier, Katerina Marazopoulou, David Arbour, and David Jensen. A sound and complete algorithm for learning causal models from relational data. *arXiv preprint arXiv:1309.6843*, 2013.

[151] Marc Maier, Katerina Marazopoulou, and David Jensen. Reasoning about independence in probabilistic models of relational data. *arXiv preprint arXiv:1302.4381*, 2013.

[152] J. Mäkelä, L. Melkas, I. Mammarella, T. Nieminen, S. Chandramouli, R. Savvides, and K. Puolamäki. Technical note: Incorporating expert domain knowledge into causal structure discovery workflows. *Biogeosciences*, 19(8):2095–2099, 2022.

[153] Ninoslav Malekovic, Maarten Vonk, Laura Birkman, Tim Sweijs, Anna Kononova, and Thomas Bäck. Angling for causality behind security: Natural causes of armed conflict in iraq. *Preprints*, February 2024.

[154] Ninoslav Malekovic, Maarten C Vonk, Laura Birkman, Tim Sweijs, Anna V Kononova, and Thomas Bäck. Applying causality to environmental security in Iraq. *Scientific Reports*, 15(1):16198, 2025.

[155] Daniel Malinsky and David Danks. Causal discovery algorithms: A practical guide. *Philosophy Compass*, 13(1):e12470, 2018.

[156] Daniel Malinsky, Ilya Shpitser, and Thomas Richardson. A potential outcomes calculus for identifying conditional path-specific effects. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pages 3080–3088. PMLR, 2019.

[157] Alexander Marx, Arthur Gretton, and Joris M Mooij. A weaker faithfulness assumption based on triple interactions. In *Proceedings of the Thirty-Seventh Conference on Uncertainty in Artificial Intelligence*, Proceedings of Machine Learning Research, pages 451–460. PMLR, 2021.

[158] Michael J. Mazarr, Arthur Chan, Alyssa Demus, Bryan Frederick, Alireza Nader, Stephanie Pezard, Julia A. Thompson, and Elina Treyger. *What Deters and Why: Exploring Requirements for Effective Deterrence of Interstate Aggression.* RAND Corporation, Santa Monica, CA, 2018.

[159] Amy McNally, Kristi Arsenault, Sujay Kumar, Shraddhanand Shukla, Pete Peterson, Shugong Wang, Chris Funk, Christa D Peters-Lidard, and James P Verdin. A land data assimilation system for sub-saharan africa food and water security applications. *Scientific data*, 4(1):1–19, 2017.

[160] Laurent Meunier and et al. Black-box optimization revisited: Improving algorithm selection wizards through massive benchmarking. *IEEE Trans. Evol. Comput.*, 26(3), 2022.

[161] Petrus Mikkola, Osvaldo A. Martin, Suyog Chandramouli, Marcelo Hartmann, Oriol Abril Pla, Owen Thomas, Henri Pesonen, Jukka Corander, Aki Vehtari, Samuel Kaski, Paul-Christian Bürkner, and Arto Klami. Prior Knowledge Elicitation: The Past, Present, and Future. *Bayesian Analysis*, pages 1 – 33, 2023.

[162] Oswaldo Morales-Nápoles and Raphaël DJM Steenbergen. Large-scale hybrid bayesian network for traffic load modeling from weigh-in-motion system data. *Journal of Bridge Engineering*, 20(1):04014059, 2015.

[163] Daniel Morato, Eduardo Berrueta, Eduardo Magaña, and Mikel Izal. Ransomware early detection by the analysis of file sharing traffic. *Journal of Network and Computer Applications*, 124:14–32, 2018.

[164] Joaquín Muñoz-Sabater, Emanuel Dutra, Anna Agustí-Panareda, Clément Albergel, Gabriele Arduini, Gianpaolo Balsamo, Souhail Boussetta, Margarita Choulga, Shaun Harrigan, Hans Hersbach, et al. Era5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth system science data*, 13(9):4349–4383, 2021.

[165] Rashelle J Musci and Elizabeth Stuart. Ensuring causal, not casual, inference. *Prevention Science*, 20:452–456, 2019.

[166] Ashley I Naimi, Stephen R Cole, and Edward H Kennedy. An introduction to g methods. *International Journal of Epidemiology*, 46(2):756–762, 12 2016.

[167] John F Nash Jr. Equilibrium points in n-person games. *Proceedings of the national academy of sciences*, 36(1):48–49, 1950.

[168] Martin Neil, Manesh Tailor, and David Marquez. Inference in hybrid bayesian networks using dynamic discretization. *Statistics and Computing*, 17(3):219–233, 2007.

[169] Farnaz Nojavan, Song S Qian, and Craig A Stow. Comparative analysis of discretization methods in bayesian networks. *Environmental Modelling & Software*, 87:64–71, 2017.

[170] R Norgaard and T Killeen. Expected utility and the truncated normal distribution. *Management Science*, 26(9):901–909, 1980.

[171] Juan Miguel Ogarrio, Peter Spirtes, and Joe Ramsey. A hybrid causal search algorithm for latent variable models. In *Proceedings of the Eighth International Conference on Probabilistic Graphical Models*, pages 368–379. PMLR, 2016.

[172] Elizabeth L Ogburn and Tyler J VanderWeele. Causal diagrams for interference. *Statistical Science*, 29(4):559–578, 2014.

[173] Rainer Opgen-Rhein and Korbinian Strimmer. From correlation to causation networks: a simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC systems biology*, 1(1):1–10, 2007.

[174] Martin J. Osborne and Ariel Rubinstein. *A course in game theory*. The MIT Press, Cambridge, USA, 1994. electronic edition.

[175] Prasad Ostwal. ostwalprasad/lgnpy: v1.0.0, 2020.

[176] Harsh Parikh, Carlos Varjao, Louise Xu, and Eric Tchetgen Tchetgen. Validating causal inference methods. In *International Conference on Machine Learning*, pages 17346–17358. PMLR, 2022.

[177] Judea Pearl. On the identification of nonparametric structural models. In Maia Berkane, editor, *Latent Variable Modeling and Applications to Causality*, pages 29–68, New York, NY, 1997. Springer New York.

[178] Judea Pearl. *Causality*. Cambridge university press, 2009.

[179] Judea Pearl and Dana Mackenzie. *The book of why: the new science of cause and effect*. Basic books, 2018.

[180] Jose M Peña. Learning acyclic directed mixed graphs from observations and interventions. In *Conference on Probabilistic Graphical Models*, volume 52, pages 392–402. PMLR, 2016.

[181] Johan Pensar, Henrik Nyman, Timo Koski, and Jukka Corander. Labeled directed acyclic graphs: a generalization of context-specific independence in directed graphical models. *Data mining and knowledge discovery*, 29(2):503–533, 2015.

[182] Emilija Perkovic. Identifying causal effects in maximally oriented partially directed acyclic graphs. In *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, volume 124 of *Proceedings of Machine Learning Research*, pages 530–539. PMLR, 03–06 Aug 2020.

[183] Franz Pernkopf, Robert Peharz, and Sebastian Tschiatschek. Introduction to probabilistic graphical models. In *Academic Press Library in Signal Processing*, volume 1, pages 989–1064. Elsevier, 2014.

[184] Jonas Peters, Dominik Janzing, and Bernhard Schölkopf. *Elements of causal inference: foundations and learning algorithms*. The MIT Press, 2017.

[185] Jonas Peters, Joris M. Mooij, Dominik Janzing, and Bernhard Schölkopf. Causal discovery with continuous additive noise models. *Journal of Machine Learning Research*, 15(58):2009–2053, 2014.

[186] M. Powell. An efficient method for finding the minimum of a function of several variables without calculating derivatives. *The Computer Journal*, 7(2), 01 1964.

[187] Clionadh Raleigh and Håvard Hegre. Population size, concentration, and civil war. a geographically disaggregated analysis. *Political geography*, 28(4):224–238, 2009.

[188] Clionadh Raleigh, Rew Linke, Håvard Hegre, and Joakim Karlsen. Introducing acled: An armed conflict location and event dataset. *Journal of peace research*, 47(5):651–660, 2010.

[189] Joseph Ramsey, Madelyn Glymour, Ruben Sanchez-Romero, and Clark Glymour. A million variables and more: the fast greedy equivalence search algorithm for learning high-dimensional graphical causal models, with an application to functional magnetic resonance images. *International Journal of Data Science and Analytics*, 3(2):121–129, 2017.

[190] J. Rapin and O. Teytaud. Nevergrad - A gradient-free optimization platform. `https://GitHub.com/FacebookResearch/Nevergrad`, 2018.

[191] Anatol Rapoport and Carol Orwant. Experimental games: A review. *Behavioral Science*, 7(1):1–37, 1962.

[192] Rafael Reuveny. Climate change-induced migration and violent conflict. *Political geography*, 26(6):656–673, 2007.

[193] Thomas Richardson and Peter Spirtes. Ancestral graph markov models. *The Annals of Statistics*, 30(4):962–1030, 2002.

[194] Thomas S Richardson. A factorization criterion for acyclic directed mixed graphs. *arXiv preprint arXiv:1406.6764*, 2014.

[195] Thomas S Richardson and James M Robins. Single world intervention graphs (swigs): A unification of the counterfactual and graphical approaches to causality. *Center for the Statistics and the Social Sciences, University of Washington Series. Working Paper*, 128(30):2013, 2013.

[196] James M Robins and Thomas S Richardson. Alternative graphical causal models and the identification of direct effects. In *Causality and Psychopathology: Finding the Determinants of Disorders and their Cures*, volume 84, pages 103–158. Oxford University Press, 2011.

[197] James M Robins, Thomas S Richardson, and Ilya Shpitser. An interventionist approach to mediation analysis. In *Probabilistic and Causal Inference: The Works of Judea Pearl*, pages 713–764, 2022.

[198] JM Robins, MA Hernán, and U Siebert. Effects of multiple interventions. *Comparative quantification of health risks: global and regional burden of disease attributable to selected major risk factors*, 1:2191–2230, 2004.

[199] Beatriz Rodríguez-Labajos and Joan Martínez-Alier. Political ecology of water conflicts. *Wiley Interdisciplinary Reviews: Water*, 2(5):537–558, 2015.

[200] Paul R Rosenbaum and Donald B Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1):41–55, 1983.

[201] Paul K Rubenstein, Sebastian Weichwald, Stephan Bongers, Joris M Mooij, Dominik Janzing, Moritz Grosse-Wentrup, and Bernhard Schölkopf. Causal consistency of structural equation models. *arXiv preprint arXiv:1707.00819*, 2017.

[202] Donald B Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Eucational Psychology*, 66(5):688–701, 1974.

[203] Donald B Rubin. Randomization analysis of experimental data: The fisher randomization test comment. *Journal of the American statistical Sssociation*, 75(371):591–593, 1980.

[204] Ariel Rubinstein. *Economic Fables*. Open Book Publishers, 1 edition, 2012.

[205] Yossi Rubner, Carlo Tomasi, and Leonidas J Guibas. The earth mover's distance as a metric for image retrieval. *International journal of computer vision*, 40(2):99, 2000.

[206] Rafael Rumí and Antonio Salmerón. Approximate probability propagation with mixtures of truncated exponentials. *International Journal of Approximate Reasoning*, 45(2):191–210, 2007.

[207] Jakob Runge. Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(7):075310, 2018.

[208] Steve Running and Q Mu. Mod16a2 modis/terra evapotranspiration 8-day l4 global 500m sin grid. *NASA LP DAAC*, 2015.

[209] Kendra Sakaguchi, Anil Varughese, and Graeme Auld. Climate wars? a systematic review of empirical analyses on the links between climate change and violent conflict. *International Studies Review*, 19(4):622–645, 2017.

[210] Igal Sason. On f-divergences: Integral representations, local behavior, and inequalities. *Entropy*, 20(5):383, 2018.

[211] Karin Schermelleh-Engel, Helfried Moosbrugger, Hans Müller, et al. Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online*, 8(2):23–74, 2003.

[212] Jacquelyn Schneider. Deterrence in and through cyberspace. *Cross-Domain Deterrence: Strategy in an Era of Complexity. Oxford University Press, Oxford*, pages 95–120, 2019.

[213] Marco Scutari. Learning bayesian networks with the bnlearn R package. *Journal of Statistical Software*, 35(3):1–22, 2010.

[214] Amartya K Sen. Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy & public affairs*, pages 317–344, 1977.

[215] David F. Shanno. Conditioning of quasi-newton methods for function minimization. *Mathematics of Computation*, 24(111):647–656, 1970.

[216] Prakash P Shenoy and James C West. Inference in hybrid bayesian networks using mixtures of polynomials. *International Journal of Approximate Reasoning*, 52(5):641–657, 2011.

[217] Eli Sherman and Ilya Shpitser. Identification and estimation of causal effects from dependent data. *Advances in Neural Information Processing Systems*, 31, 2018.

[218] Shohei Shimizu, Patrik O Hoyer, Aapo Hyvärinen, Antti Kerminen, and Michael Jordan. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research*, 7(72):2003–2030, 2006.

[219] Ilya Shpitser. Segregated graphs and marginals of chain graph models. *Advances in Neural Information Processing Systems*, 28, 2015.

[220] Ilya Shpitser and Judea Pearl. Identification of joint interventional distributions in recursive semi-markovian causal models. In *Proceedings of the 21st National Conference on Artificial Intelligence - Volume 2*, AAAI'06, page 1219–1226. AAAI Press, 2006.

[221] Ilya Shpitser, Thomas S. Richardson, and James M. Robins. *Multivariate Counterfactual Systems and Causal Graphical Models*, page 813–852. Association for Computing Machinery, New York, NY, USA, 1 edition, 2022.

[222] Ilya Shpitser and Eric Tchetgen Tchetgen. Causal inference with a graphical hierarchy of interventions. *Annals of Statistics*, 44(6):2433–2466, 2016.

[223] Martin Shubik. Eminent paper series the present and future of game theory. *The Singapore Economic Review*, 57(01):1250001, 2012.

[224] Ricardo Silva. Observational-interventional priors for dose-response learning. *Advances in Neural Information Processing Systems*, 29, 2016.

[225] Rune T Slettebak. Don't blame the weather! climate-related natural disasters and civil conflict. *Journal of Peace Research*, 49(1):163–176, 2012.

[226] Jim Q Smith and Paul E Anderson. Conditional independence and chain event graphs. *Artificial Intelligence*, 172(1):42–68, 2008.

[227] Mauricio Gonzalez Soto, David Danks, Hugo J Escalante Balderas, and L Enrique Sucar. Choosing with unknown causal information: Action-outcome probabilities for decision making can be grounded in causal models. *arXiv preprint arXiv:1907.11752*, 2019.

[228] David J Spiegelhalter and Steffen L Lauritzen. Sequential updating of conditional probabilities on directed graphical structures. *Networks*, 20(5):579–605, 1990.

[229] Peter Spirtes and Clark Glymour. An algorithm for fast recovery of sparse causal graphs. *Social Science Computer Review*, 9(1):62–72, 1991.

[230] Peter Spirtes, Clark N. Glymour, and Richard Scheines. Causality from probability. In *Conference Proceedings: Advanced Computing for the Social Sciences*, 1990.

[231] Peter Spirtes, Clark N Glymour, Richard Scheines, and David Heckerman. *Causation, prediction, and search*. MIT press, 2nd edition, 2000.

[232] Global Spatially-Disaggregated Crop Production Statistics. Data for 2010 version 2.0. *IFPRI: Washington, DC, USA*, 2019.

[233] Niki van Stein, Elena Raponi, Zahra Sadeghi, Niek Bouman, Roeland C. H. J. Van Ham, and Thomas Bäck. A comparison of global sensitivity analysis methods for explainable ai with an application in genomic prediction. *IEEE Access*, 10:103364–103381, 2022.

[234] Pasquale Stirparo, David Bizeul, Brian Bell, Ziv Chang, Joel Esler, Kristopher Bleich, Maite Moreno, J Monnappa KA, Paul Hutchinson Capmany, Boris Ivanov, et al. Apt groups and operations, 2019.

[235] R. Storn and K. Price. Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4):341–359, 1997.

[236] Luis Enrique Sucar. *Probabilistic Graphical Models: Principles and Applications*. Springer Nature, 2020.

[237] Julia E Sullivan and Dmitriy Kamensky. How cyber-attacks in ukraine show the vulnerability of the u.s. power grid. *The Electricity Journal*, 30(3):30–35, 2017.

[238] Ralph Sundberg and Erik Melander. Introducing the ucdp georeferenced event dataset. *Journal of peace research*, 50(4):523–532, 2013.

[239] Eric J Tchetgen Tchetgen and Tyler J VanderWeele. On causal inference in the presence of interference. *Statistical Methods in Medical Research*, 21(1):55–75, 2012.

[240] Santtu Tikka, Antti Hyttinen, and Juha Karvanen. Identifying causal effects via context-specific independence relations. *Advances in Neural Information Processing Systems*, 32, 2019.

[241] Jaroslav Tir and Paul F Diehl. Demographic pressure and interstate conflict: linking population growth and density to militarized disputes and wars, 1930-89. *Journal of Peace Research*, 35(3):319–339, 1998.

[242] Gregory F Treverton, Andrew Thvedt, Alicia R Chen, Kathy Lee, and Madeline McCue. Addressing hybrid threats, 2018.

[243] Brandon Valeriano and Ryan C Maness. The dynamics of cyber conflict between rival antagonists, 2001–11 (dataset updated 2022). *Journal of Peace Research*, 51(3):347–360, 2022.

[244] Benito van der Zander, Maciej Liśkiewicz, and Johannes Textor. Separators and adjustment sets in causal graphs: Complete criteria and an algorithmic framework. *Artificial Intelligence*, 270:1–40, 2019.

[245] Tyler J VanderWeele and Miguel A Hernan. Causal inference under multiple versions of treatment. *Journal of Causal Inference*, 1(1):1–20, 2013.

[246] Diederick Vermetten, Jacob de Nobel, Furong Ye, Hao Wang, Carola Doerr, and Thomas Bäck. IOHprofiler: Iterative Optimization Heuristics Profiler. `https://github.com/IOHprofiler`, 2018. wiki available at `https://iohprofiler.github.io/`.

[247] Claudia Vitolo, Marco Scutari, Mohamed Ghalaieny, Allan Tucker, and Andrew Russell. Modeling air pollution, climate, and health data using bayesian networks: A case study of the english regions. *Earth and Space Science*, 5(4):76–88, 2018.

[248] John von Neumann and Oskar Morgenstern. *Theory of Games and Economic Behavior*. Princeton University Press, Princeton, 2004.

[249] Maarten C Vonk, Sebastiaan Brand, Ninoslav Malekovic, Thomas Bäck, Alfons Laarman, and Anna V Kononova. Balancing computational cost and accuracy in inference of continuous bayesian networks. In *International Conference on Probabilistic Graphical Models*, pages 361–381. PMLR, 2024.

[250] Maarten C Vonk, Anna V Kononova, Thomas Bäck, and Tim Sweijs. Multi-agent influence diagrams to hybrid threat modeling. *The Journal of Defense Modeling and Simulation*, 0(0), 2025.

[251] Maarten C Vonk, Ninoslav Malekovic, Thomas Bäck, and Anna V Kononova. Disentangling causality: assumptions in causal discovery and inference. *Artificial Intelligence Review*, 56(9):10613–10649, 2023.

[252] Maarten C Vonk, Mauricio Gonzalez Soto, and Anna V Kononova. Graphical models for decision-making: Integrating causality and game theory. *arXiv preprint arXiv:2504.13210*, 2025.

[253] Maarten C Vonk, Diederick Vermetten, Jacob de Nobel, Sebastiaan Brand, Ninoslav Malekovic, Thomas Bäck, Alfons Laarman, and Anna V Kononova. Optimizing causal interventions in hybrid bayesian networks. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 245–256. Springer, 2024.

[254] Peter Wakker and Daniel Deneffe. Eliciting von neumann-morgenstern utilities when probabilities are distorted or unknown. *Management science*, 42(8):1131–1150, 1996.

[255] Junqi Wang, Pei Wang, and Patrick Shafto. Efficient discretizations of optimal transport. *arXiv preprint arXiv:2102.07956*, 2021.

[256] Yi Wang, Yuan Sun, Ji-Ying Li, and Sun-Tao Xia. Air defense threat assessment based on dynamic bayesian network. In *2012 International Conference on Systems and Informatics (ICSAI2012)*, pages 721–724. IEEE, 2012.

[257] Michael P Wellman, Karl Tuyls, and Amy Greenwald. Empirical game-theoretic analysis: A survey. *arXiv preprint arXiv:2403.04018*, 2024.

[258] Mikael Wigell. Democratic deterrence: How to dissuade hybrid interference. *The Washington Quarterly*, 44(1):49–67, 2021.

[259] Lingbo B Xiao, Xiuqi Q Fang, and Yu Ye. Reclamation and revolt: Social responses in eastern inner mongolia to flood/drought-induced refugees from the north china plain 1644–1911. *Journal of Arid Environments*, 88:9–16, 2013.

[260] Guanhua Yan, Ritchie Lee, Alex Kent, and David Wolpert. Towards a bayesian network game framework for evaluating ddos attacks and defense. In *Proceedings of the 2012 ACM conference on Computer and communications security*, pages 553–566, 2012.

[261] Weiwen Yin. Climate shocks, political institutions, and nomadic invasions in early modern east asia. *Journal of Conflict Resolution*, 64(6):1043–1069, 2020.

[262] Jessica G Young, Miguel A Hernán, and James M Robins. Identification, estimation and approximation of risk under interventions that depend on the natural value of treatment using observational data. *Epidemiologic Methods*, 3(1):1–19, 2014.

[263] Jiji Zhang. A comparison of three occam's razors for markovian causal models. *The British Journal for the Philosophy of Science*, 64:423–448, 2013.

[264] Jiji Zhang and Peter Spirtes. Intervention, determinism, and the causal minimality condition. *Synthese*, 182(3):335–347, 2011.

[265] Jiji Zhang and Peter Spirtes. The three faces of faithfulness. *Synthese*, 193(4):1011–1027, 2015.

[266] Kun Zhang and Aapo Hyvärinen. On the identifiability of the post-nonlinear causal model. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, page 647–655, 2009.

[267] Angela Yaqian Zhu, Nandita Mitra, and Jason Roy. Addressing positivity violations in causal effect estimation using gaussian process priors. *Statistics in Medicine*, 42(1):33–51, 2023.

[268] Jinlin Zhu, Guohao Peng, and Danwei Wang. Dual-domain-based adversarial defense with conditional vae and bayesian network. *IEEE Transactions on Industrial Informatics*, 17(1):596–605, 2020.

Summary

Effective policy-making requires a clear understanding of what truly causes a problem. Only through such understanding can policymakers develop targeted interventions that achieve the desired outcomes. However, establishing causal relationships presents a significant challenge, as observed patterns do not automatically indicate true cause-and-effect relationships.

This thesis provides various tools for applying causal analysis methods to complex security environments. What distinguishes some of these methods is their consideration of how different parties react to one another. In complex security environments, countries, organizations, and groups do not act in isolation; they anticipate each other's actions and adapt their behavior accordingly.

The thesis presents a structured framework that supports policymakers in selecting appropriate analytical methods for specific policy questions. For each method, the necessary assumptions are explained along with practical guidance for implementation. Additionally, it demonstrates how causal analysis can be combined with strategic thinking by incorporating the intentions and potential reactions of adversaries.

A concrete methodological innovation involves the development of an automated method for identifying optimal intervention points. This approach enables policymakers to specify desired policy objectives, after which algorithms determine which interventions are most effective and how relevant variables should be adjusted.

The practical applicability is demonstrated through two current security challenges: hybrid threats and climate-related conflicts. These cases are characterized by inherent uncertainty, multiple strategic actors, and mutual dependencies.

The research results in an analytical toolkit for decision-making in complex security environments. These instruments not only facilitate more informed policy choices but also contribute to developing the analytical capabilities necessary for navigating strategic complexity in contemporary security issues.

Samenvatting

Bij het maken van beleid is het cruciaal om te begrijpen wat werkelijk de oorzaak is van een probleem. Alleen zo kunnen beleidsmakers doeltreffende maatregelen nemen die het gewenste effect sorteren. Het vaststellen van causale relaties vormt echter een uitdaging, aangezien waargenomen patronen niet automatisch wijzen op duidelijke oorzaak-gevolg relaties.

Dit proefschrift biedt verschillende handvaten bij het toepassen van causale analysemethoden op complexe veiligheidsvraagstukken, waarbij specifiek aandacht wordt besteed aan methoden die rekening houden met hoe verschillende partijen op elkaar reageren. In veiligheidssituaties handelen landen, organisaties of groepen immers niet in isolatie, maar ze anticiperen op elkaars acties en passen hun gedrag daarop aan.

Het proefschrift presenteert een gestructureerd raamwerk dat beleidsmakers ondersteunt bij het selecteren van geschikte analysemethoden voor specifieke beleidsvragen. Voor elke methode wordt uitgelegd welke aannames je moet maken en hoe je deze in de praktijk toepast. Daarnaast wordt getoond hoe je causale analyse kunt combineren met strategisch denken door de intenties en mogelijke reacties van tegenstanders mee te nemen.

Een concrete methodologische innovatie betreft de ontwikkeling van een geautomatiseerde methode voor het identificeren van optimale interventiepunten. Deze benadering stelt beleidsmakers in staat om gewenste beleidsdoelen te specificeren, waarna algoritmisch wordt bepaald welke interventies het meest effectief zijn en op welke wijze relevante variabelen moeten worden aangepast.

De praktische toepasbaarheid wordt gedemonstreerd aan de hand van twee actuele veiligheidsuitdagingen: hybride dreigingen en klimaatgerelateerde conflicten. Deze casussen worden gekenmerkt door inherente onzekerheid, meerdere strategisch handelende actoren en wederzijdse afhankelijkheden.

Het onderzoek resulteert in een analytische gereedschapskist voor besluitvorming

in complexe veiligheidsomgevingen. Deze instrumenten faciliteren niet alleen meer onderbouwde beleidskeuzes, maar dragen tevens bij aan de ontwikkeling van analytische capaciteiten die noodzakelijk zijn voor het adresseren van strategische complexiteit in hedendaagse veiligheidsvraagstukken.

Acknowledgements

I wish to express my sincere gratitude to all who have supported me over the past four years. I am particularly grateful to The Hague Centre for Strategic Studies for granting me the opportunity to undertake my PhD at the intersection of computer science and strategic studies. This research was supported by the Ministry of Foreign Affairs and the Ministry of Defence of the Netherlands and I am grateful for their interest in and commitment to advancing research on the development of innovative tools for strategic studies.

Second, I am deeply grateful to my supervisors, Prof. Dr. Thomas Bäck, Dr. Anna Kononova, and Dr. Tim Sweijs. I thank Prof. Dr. Bäck for his oversight of the entire project and his steady guidance throughout. I am especially grateful to Dr. Kononova for her consistent support and engagement in the day-to-day process of doing rigorous research. I also thank Dr. Sweijs for enriching my work with his perspective from applied strategic studies, which broadened the scope and relevance of this dissertation.

Third, I would like to thank all my colleagues at The Hague Centre for Strategic Studies for their support, encouragement, and intellectual engagement throughout my research journey. I am particularly grateful for the opportunity to collaborate on projects that bridged academic research and policy practice, and for the many insightful discussions that helped shape and refine the ideas developed in this dissertation.

Fourth, I would like to thank my colleagues at the Natural Computing Group and, more broadly, at the Leiden Institute of Advanced Computer Science. I am especially grateful to my direct collaborators, Sebastiaan Brand, Dr. Alfons Laarman, Jacob de Nobel, and Dr. Diederick Vermetten, for their valuable expertise in decision diagrams and optimization. I also wish to thank my collaborators in Mexico, in particular Dr. Enrique Sucar and Dr. Julio César Muñoz Benítez, for warmly welcoming me to the National Institute of Astrophysics, Optics and Electronics in Cholula, and Mauricio

González Soto for his dedicated work with me on causal game theory.

Finally, I wish to express my heartfelt gratitude to my girlfriend, family, and friends for their unwavering support throughout this journey. In particular, I would like to thank my uncle, Aad Vonk, whose enthusiasm for scientific research has been a constant motivator throughout my PhD journey.

About the Author

Maarten Vonk was born on September 12, 1993, in Rotterdam. After completing his Bachelor's degree in mathematics at the University of Amsterdam in 2016, Maarten continued his studies and obtained a master's degree in applied mathematics with a specialization in discrete optimization from the Technical University of Delft in 2018. Before starting his PhD, he gained some working experience as a data scientist and software consultant at Transavia and Ortec Finance, respectively. In June 2021, Maarten started working as a data scientist at the Hague Centre for Strategic Studies. Besides his work as a data scientist, he was funded by a joint research fund of the Dutch Ministry of Foreign Affairs and the Dutch Ministry of Defence, called Progress. Because of this fund, Maarten was able to pursue a PhD to develop quantitative tools for policy making as a guest PhD candidate at the Leiden Institute of Advanced Computer Science under the supervision of Prof. dr. Thomas Bäck, Dr. Anna V. Kononova and Dr. Tim Sweijs. He conducted research as part of Dr. Kononova's Efficient Heuristic Optimisation (EcHO) group. During his PhD studies, he took a course in scientific conduct.