Conditional Independence for Causal Reasoning

A. Philip Dawid

ECSQARU
Utrecht
8 July 2013

SEEING AND DOING

- Causality is about the effects of *interventions* (doing)
- To discover these we really should *experiment*
- If we can’t, is there anything sensible we can conclude from observational data (seeing)?
- No amount of clever analysis of purely observational data can replace experimentation
  - we have to make unverifiable assumptions

SEEING

- Association
  - describe stochastic dependence and independence
- Conditional Independence (CI)

$$X \perp\!\!\!\perp Y \mid Z \; [P]$$

  - attribute of a joint probability distribution *P*

$$\begin{aligned} p(x, y \mid z) &= p(x \mid z)\, p(y \mid z) \\ p(x \mid y, z) &= p(x \mid z) \end{aligned}$$

Properties of CI

P1 $X \perp\!\!\!\perp Y \mid Z \quad \Rightarrow \quad Y \perp\!\!\!\perp X \mid Z$

P2 $X \perp\!\!\!\perp Y \mid X$

P3 $X \perp\!\!\!\perp Y \mid Z, \quad W \leq Y \quad \Rightarrow \quad X \perp\!\!\!\perp W \mid Z$

P4 $X \perp\!\!\!\perp Y \mid Z, \quad W \leq Y \quad \Rightarrow \quad X \perp\!\!\!\perp Y \mid (W, Z)$

P5
$$\left.\begin{array}{l} X \perp\!\!\!\perp Y \mid Z \\ \text{and} \\ X \perp\!\!\!\perp W \mid (Y, Z) \end{array}\right\} \quad \Rightarrow \quad X \perp\!\!\!\perp (Y, W) \mid Z$$

Algebraic Representation

- We can make these properties the *axioms* of a formal algebraic theory
  - *separoid*
  - *semi-graphoid*
- Other applications too
- Can use to represent and manipulate CI without referring back to *P*
- Not complete

Use As Axioms

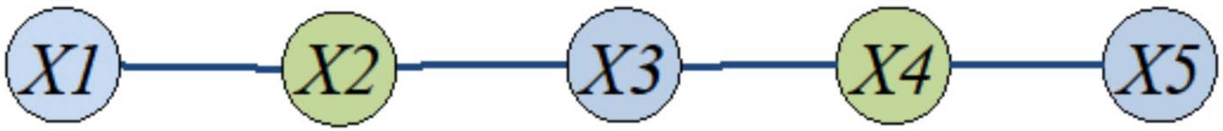
Suppose:

(i). $X_3 \perp\!\!\!\perp X_1 \mid X_2$

(ii). $X_4 \perp\!\!\!\perp (X_1, X_2) \mid X_3$

(iii). $X_5 \perp\!\!\!\perp (X_1, X_2, X_3) \mid X_4$

Then $X_3 \perp\!\!\!\perp (X_1, X_5) \mid (X_2, X_4)$.



Proof

Applying P4 and P1 in turn to (ii), we obtain

$$X_1 \perp\!\!\!\perp X_4 \mid (X_2, X_3), \tag{1}$$

while from (i) and P1 we have

$$X_1 \perp\!\!\!\perp X_3 \mid X_2. \tag{2}$$

On applying P5 to (2) and (1), we now deduce

$$X_1 \perp\!\!\!\perp (X_3, X_4) \mid X_2 \tag{3}$$

whence, by P4 and P1,

$$X_3 \perp\!\!\!\perp X_1 \mid (X_2, X_4). \tag{4}$$

Also, by (iii) and P4 we have

$$X_5 \perp\!\!\!\perp (X_1, X_3) \mid (X_2, X_4) \tag{5}$$

and so, by P4 and P1,

$$X_3 \perp\!\!\!\perp X_5 \mid (X_1, X_2, X_4). \tag{6}$$

The result now follows on applying P5 to (4) and (6).

Graphical Representation

- Certain collections of CI properties can be described and manipulated using a Directed Acyclic Graph representation
  - *very far from complete*
- Each CI property is *represented* by a graphical separation property
  - *d-separation*
  - *moralization*

DAG construction

Given a distribution over ordered set of variables

$$V^N = (V_1, \ldots, V_N),$$

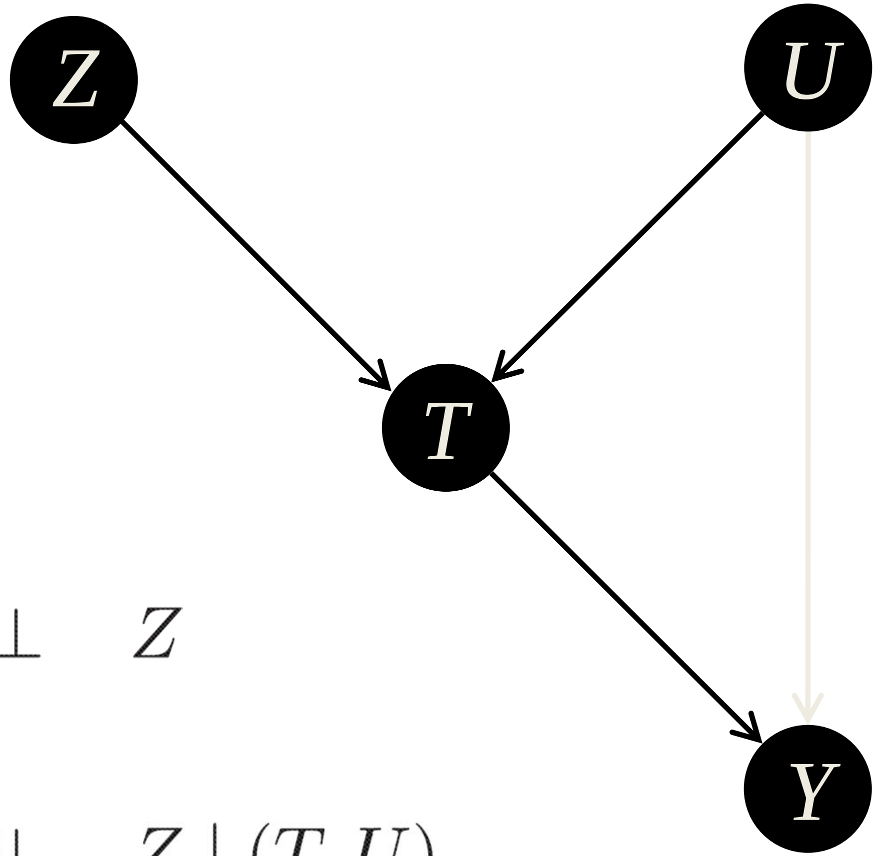
Construct DAG with the (V_i) as vertices as follows:

For $i = 0, \ldots, N-1$:

- $S_i =$ subset of V^i such that $C_{i+1} : V_{i+1} \perp\!\!\!\perp V^i \mid S_i$
- Insert an arrow from each $V_j \in S_i$ into V_{i+1}

Resulting DAG represents exactly those CI properties algebraically deducible from $C_1, \ldots, C_N$

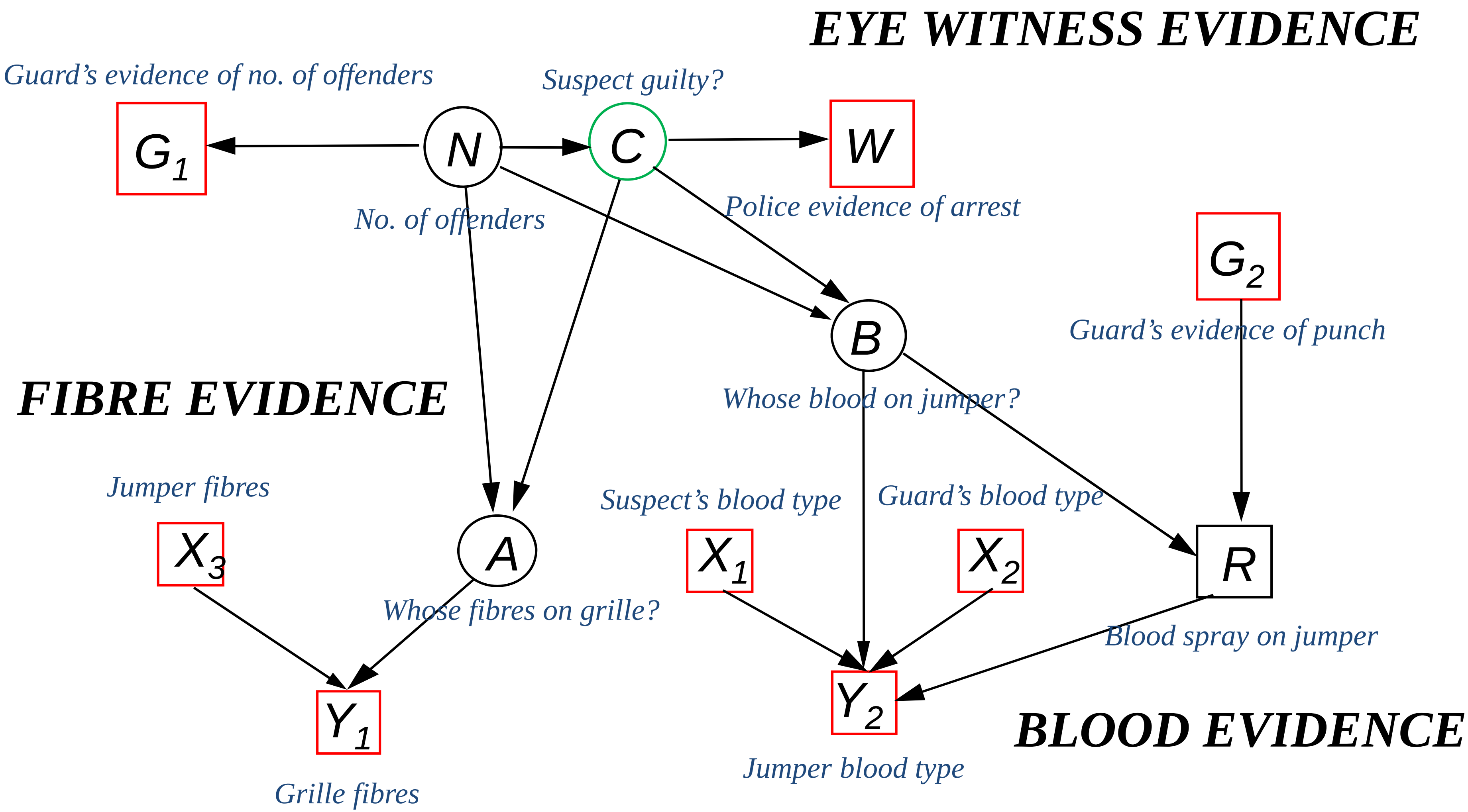
Example



$$U \perp\!\!\!\perp Z$$

$$Y \perp\!\!\!\perp Z \mid (T, U)$$

Criminal Evidence



Criminal Evidence

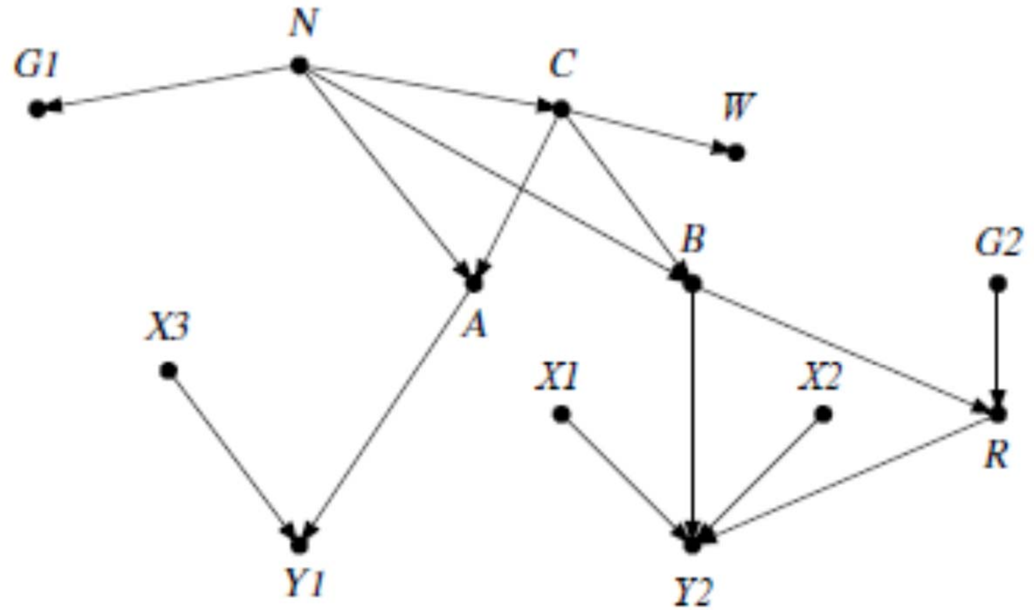


Figure 6.1: Directed graph $\mathcal{D}$ for criminal evidence

Query

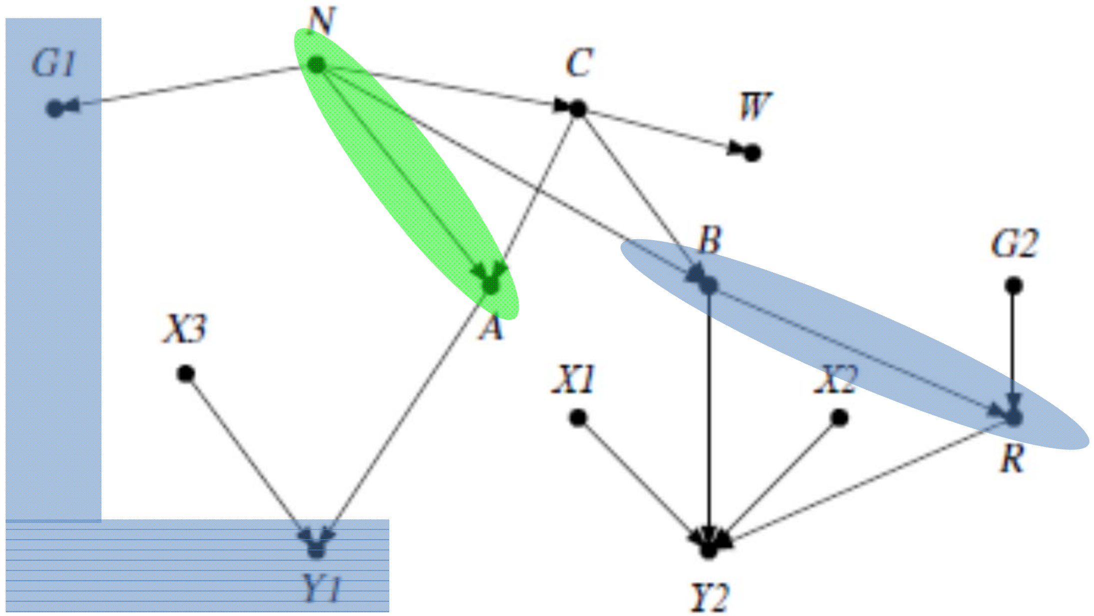


Figure 6.1: Directed graph $\mathcal{D}$ for criminal evidence

$$(B, R) \perp\!\!\!\perp (G1, Y1) \mid (A, N) \text{ ??}$$

Ancestral Graph

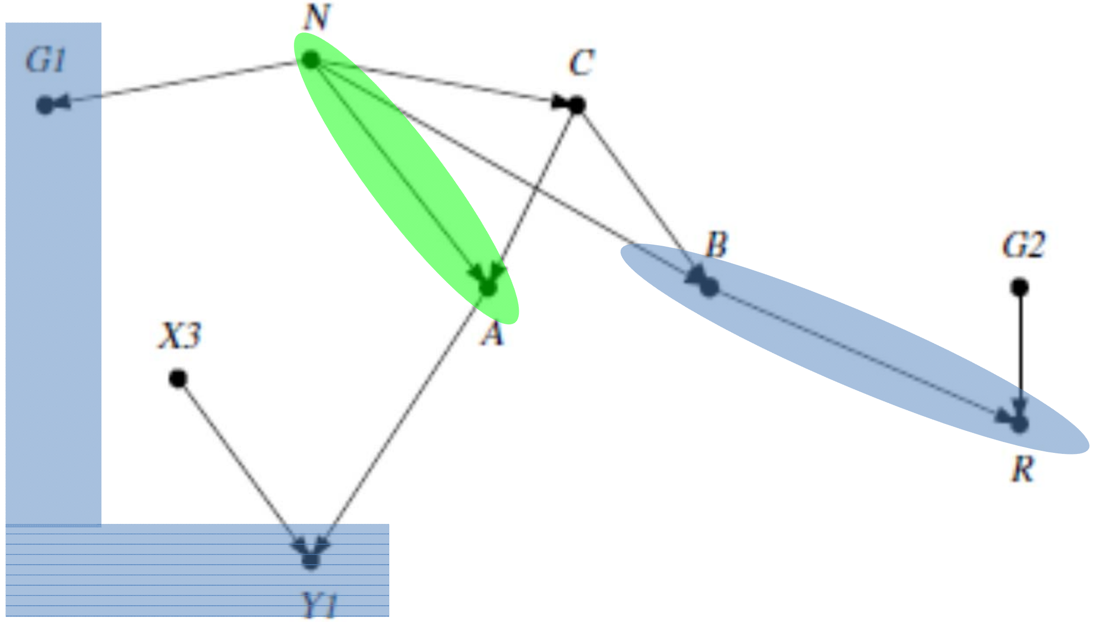


Figure 6.2: Ancestral subgraph $\mathcal{D}'$

$$(B, R) \perp\!\!\!\perp (G1, Y1) \mid (A, N) \; ??$$

Moralization

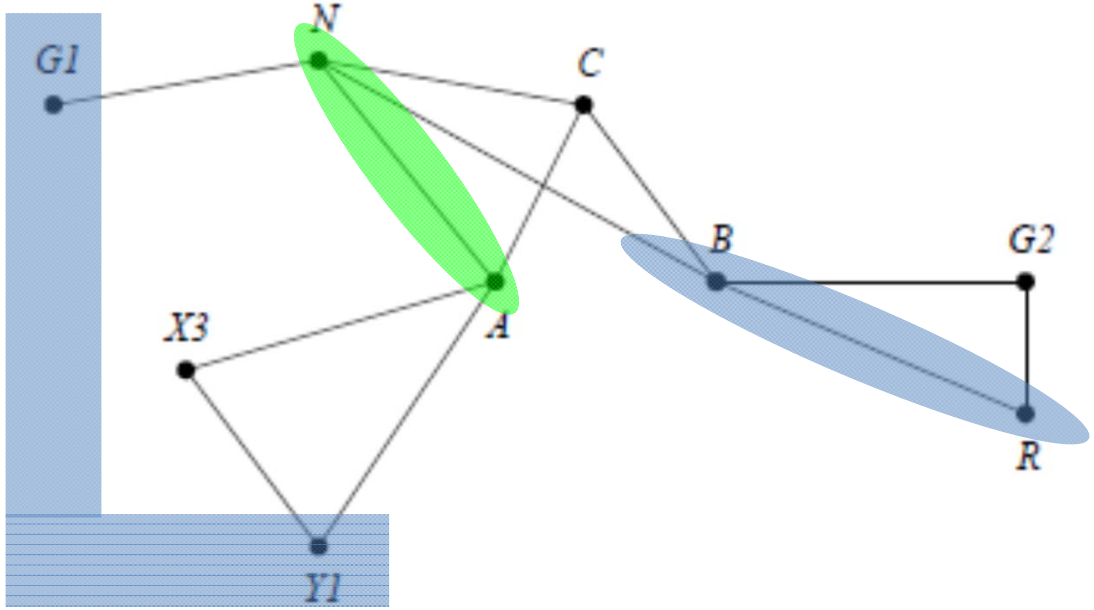
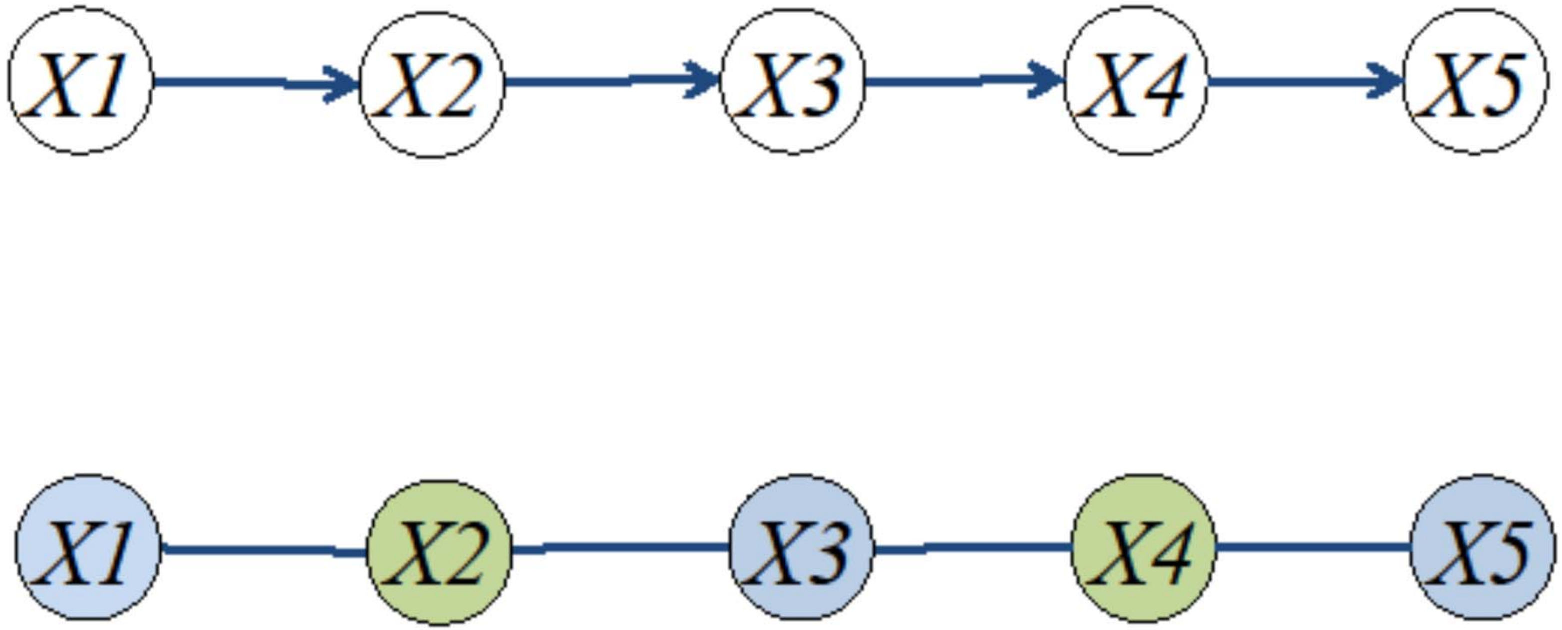


Figure 6.3: Moralized ancestral subgraph $\mathcal{G}'$

$$(B, R) \perp\!\!\!\perp (G1, Y1) \mid (A, N)$$

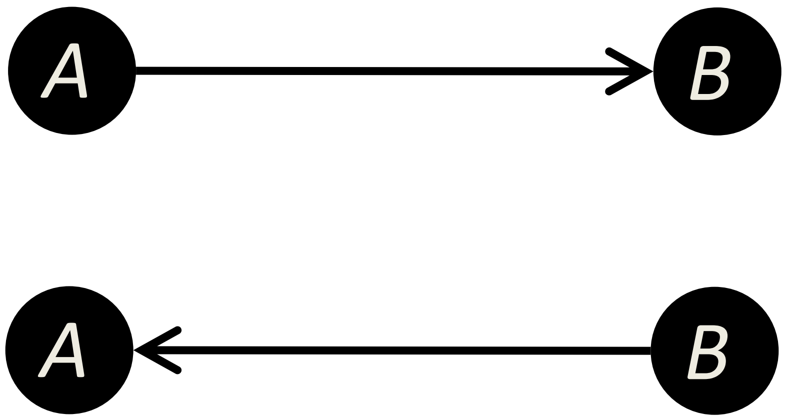
Markov Chain



$$X_3 \perp\!\!\!\perp (X_1, X_5) \mid (X_2, X_4)$$

Markov Equivalence

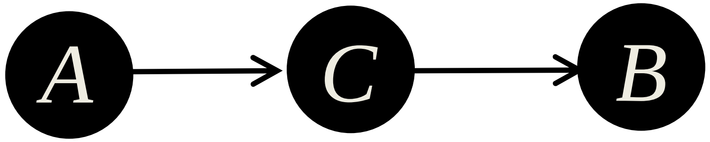
Same skeleton and immoralities



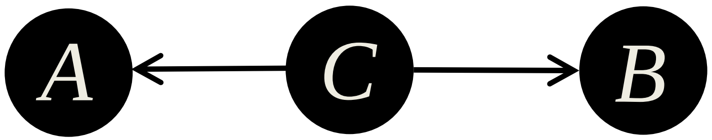
No structure

Markov Equivalence

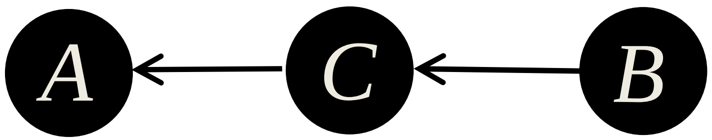
Same skeleton and immoralities



$$A \perp\!\!\!\perp B \mid C$$

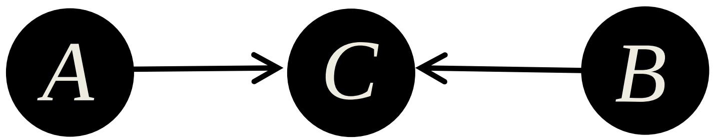


$$A \perp\!\!\!\perp B \mid C$$



$$A \perp\!\!\!\perp B \mid C$$

Markov Equivalence



$$A \perp\!\!\!\perp B$$

$$A \perp\!\!\!\perp B$$

Points to Remember

- The DAG is *nothing but* an indirect way of describing a set of CI relationships
- Clear semantics (moralization)
- May be several representations, or none
- ***Arrows have no intrinsic meaning***
  - ***CI is non-directional!***
- Represented relationships unaffected by other unmentioned, omitted variables,…
- ***Nothing to do with causality…***

DOING

- However, DAGs are often used to represent causal relationships
- But the semantics of such a representation are typically informal, ambiguous, unclear…

“Reification” of a DAG

- (Some) arrows represent direction of influence, direct cause,…
- (Some) directed paths represent “causal pathways”
- If these exist in all equivalent DAG representations they are “truly causal”

What do above causal terms mean?
Why/how do they relate to DAGs?

Probabilistic Causality

- *Weak Causal Markov* assumption:
  - If *X* and *Y* have no common cause (including each other), they are probabilistically independent
- *Causal Markov* assumption:
  - A variable is probabilistically independent of its non-effects, given its direct causes

What do above causal terms mean?

When/how widely do these assumptions hold?

Causal DAG

A *causal DAG* is a DAG in which:

1) the lack of an arrow from V_j to V_m can be interpreted as the absence of a direct causal effect of V_j on V_m (relative to the other variables on the graph)
2) all common causes (even if unmeasured) of any pair of variables on the graph are themselves on the graph

Then Causal Markov $\Rightarrow$ Markov

Converse???

Some problems

- Multiple interpretations of the same object (DAG)
  - ambiguous and confusing
- Causal interpretation informal and obscure
  - which comes first, the process or the DAG?

➢ We need a clear formal language, with explicit semantics, by which we can describe and manipulate causal properties

➢ This should not commit us to any particular causal assumptions

Causality and Intervention

- Causality = response of a system to an (actual or proposed) intervention
- Typically we can only observe undisturbed ("idle") system
- Causal inference will requires *assumptions* relating idle and interventional regimes
- Want a language to express such assumptions

Intervention Variables

Variable F_X describing *kind* of intervention at X

$F_X = x$: *manipulate* X to value x

$F_X = \emptyset$: hands off!

Different settings of intervention variables determine different joint distributions (so parameter, not random, variables)

Assume $F_X = x \Rightarrow X = x$ (can relax…)

— no other hard-and-fast assumptions

A *Possible* Assumption: Modularity

- "*A* causes *B*"
- Knowing value *a* taken by *A*, do not need to know HOW this arose (by intervention, or naturally) in order to predict *B*
- Conditional distribution of *B* given *A* is a modular component, transferable across regimes
- $p(B \mid A, F_A)$ does not depend on F_A
  - $B \perp\!\!\!\perp F_A \mid A$

Extended Conditional Independence

- Such "extended CI" properties can be formally manipulated using the same algebraic rules as for regular CI

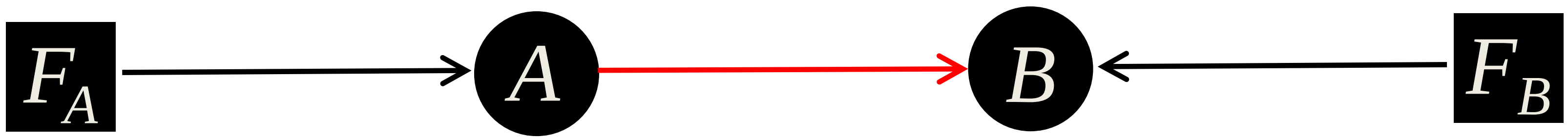
- Allows us to determine consequences of our input assumptions

  - *Causal inference*

Augmented DAG

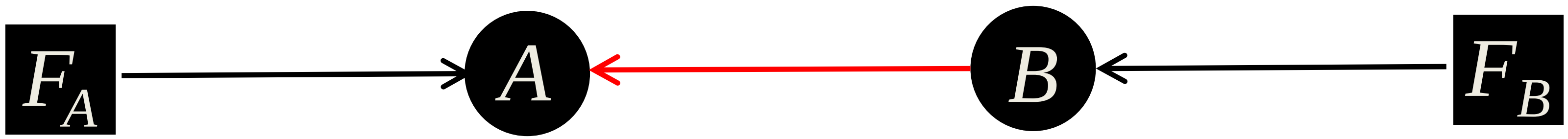
- Include intervention indicators in DAG
- Explicit causal interpretation
  - using *moralization* to express ECI
  - causality *NOT* (directly) represented by arrows

Making sense of the arrows



$$B \perp\!\!\!\perp F_A \,|\, A \;(\& F_B)$$

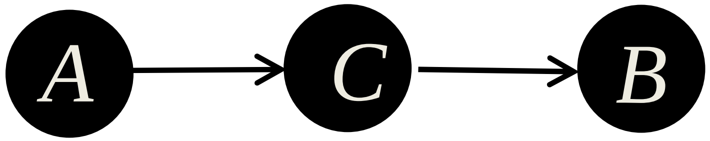
"*A* causes *B*"



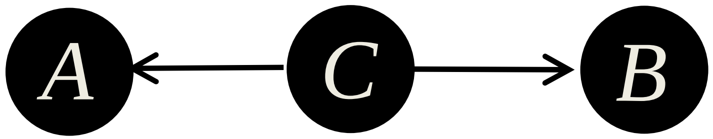
$$A \perp\!\!\!\perp F_B \,|\, B \;(\& F_A)$$

"*B* causes *A*"

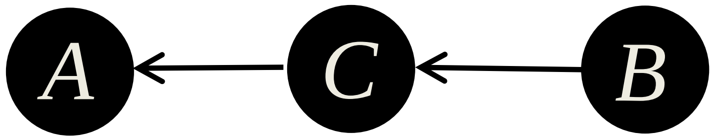
Markov Equivalence



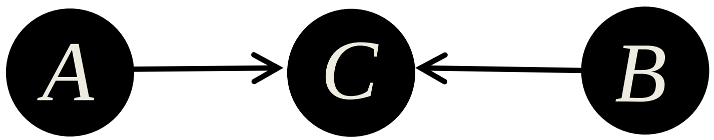
$$A \perp\!\!\!\perp B \mid C$$



$$A \perp\!\!\!\perp B \mid C$$



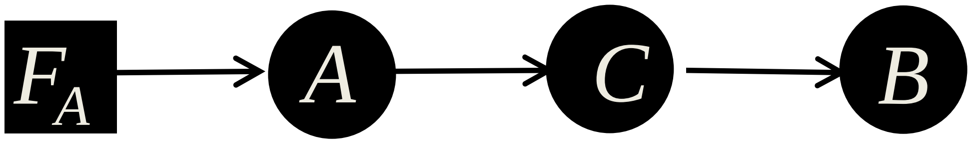
$$A \perp\!\!\!\perp B \mid C$$



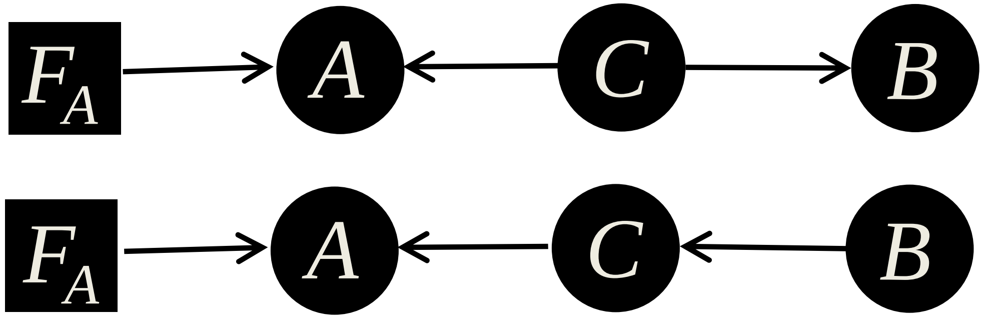
$$A \perp\!\!\!\perp B$$

$$A \perp\!\!\!\perp B$$

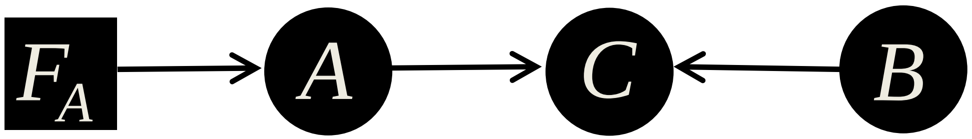
Markov Non-Equivalence



$$\begin{cases} B \perp\!\!\!\perp (A, F_A) \mid C \\ C \perp\!\!\!\perp F_A \mid A \end{cases}$$



$$\begin{cases} B \perp\!\!\!\perp (A, F_A) \mid C \\ C \perp\!\!\!\perp F_A \end{cases}$$



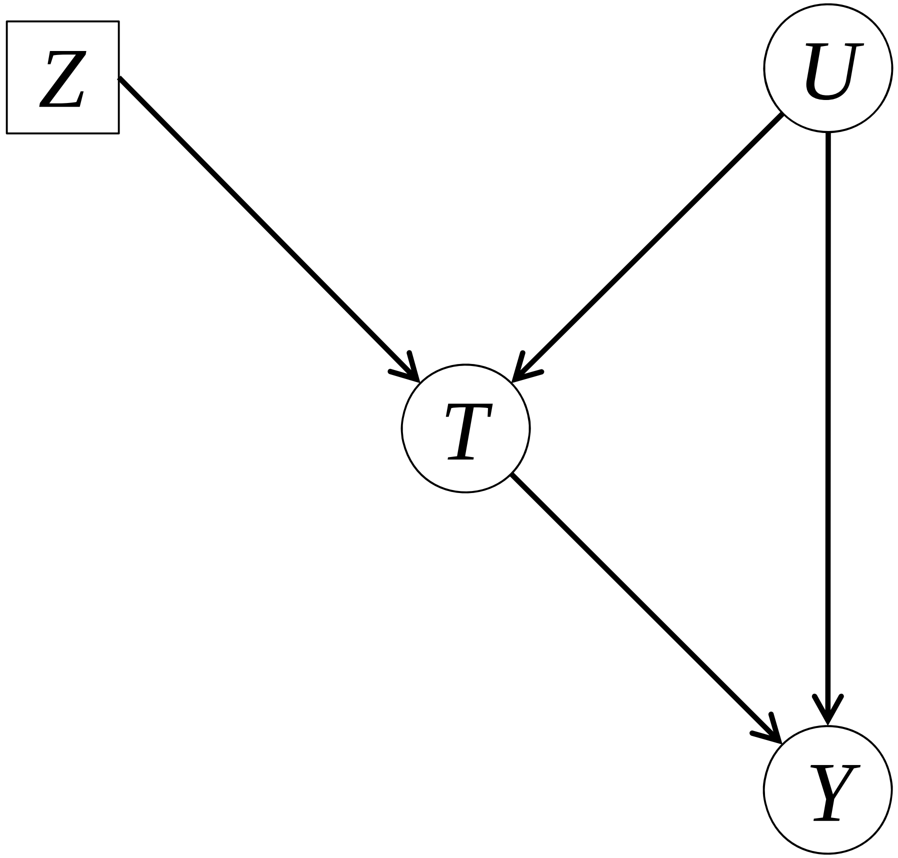
$$\begin{cases} B \perp\!\!\!\perp (A, F_A) \\ C \perp\!\!\!\perp F_A \mid (A, B) \end{cases}$$

$$\left\{ B \perp\!\!\!\perp (A, F_A) \right.$$

Pearlian DAG

- Pearl's interpretation of a DAG as causal:
  - implicit addition of an intervention node for each random node
- Relates regimes that intervene on any set of variables (or none)
- When valid, allows causal inference from observational data

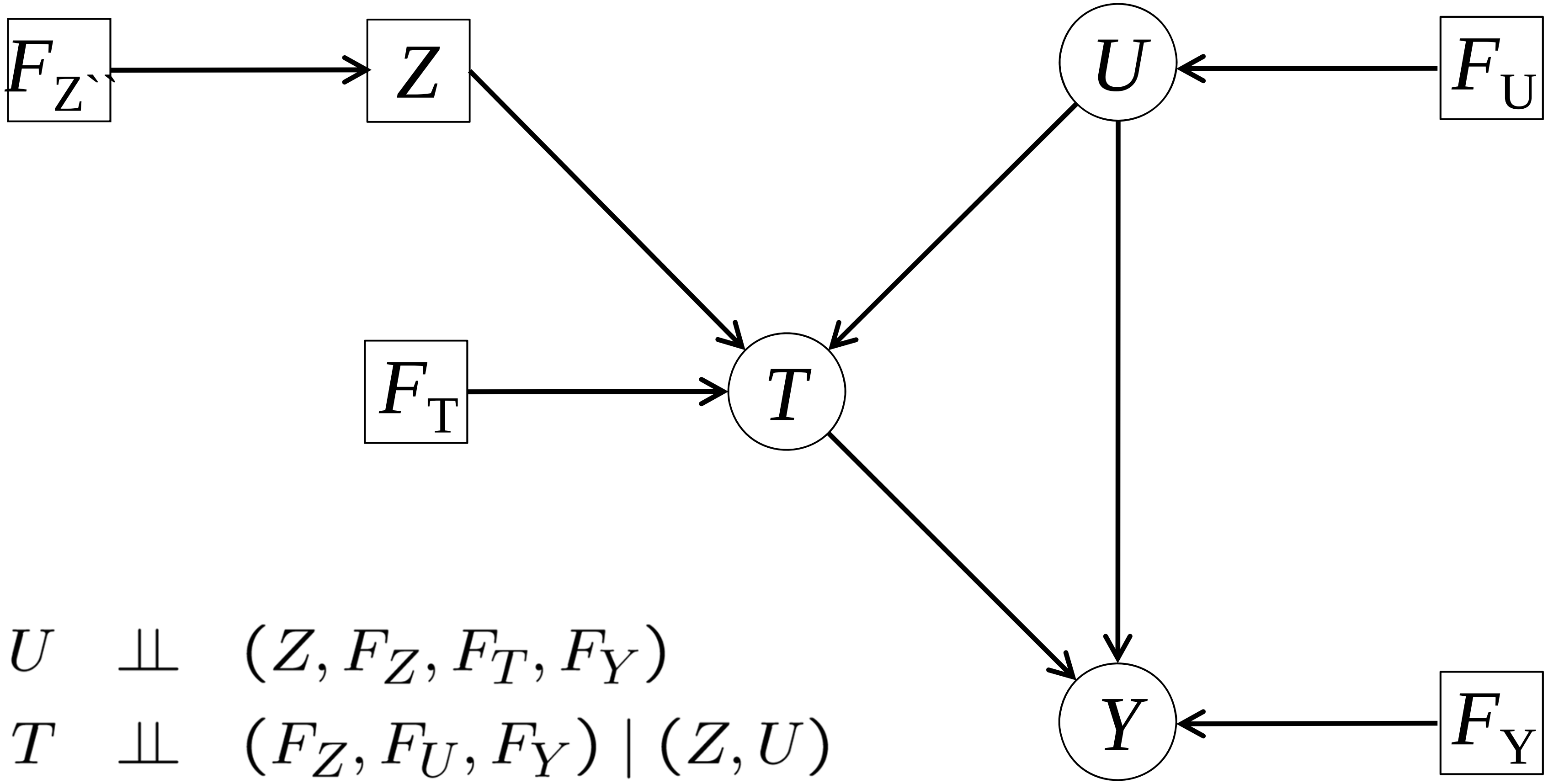
Pearlian DAG



$$U \perp\!\!\!\perp Z$$

$$Y \perp\!\!\!\perp Z \mid (T, U)$$

Intervention DAG



$$U \perp\!\!\!\perp (Z, F_Z, F_T, F_Y)$$
$$T \perp\!\!\!\perp (F_Z, F_U, F_Y) \mid (Z, U)$$
$$Y \perp\!\!\!\perp (Z, F_Z, F_T, F_U) \mid (T, U)$$

Can we just add intervention variables to a DAG?

- Behaviour of system when kicked need not bear any relationship to its behaviour when observed
- If $A \perp\!\!\!\perp B \quad (A \perp\!\!\!\perp B \mid \text{ancestors}),$

  on adding interventions, neither of *A* nor *B* can cause the other (converse of weak causal Markov property??)
  - why need this be?

“Causal Discovery”

- Gather observational data on system
- Infer *conditional independence properties* of joint distribution
- Fit a *DIRECTED ACYCLIC GRAPH* model to represent these
- Interpret this model *CAUSALLY*

OK???

- When is this meaningful?
- What does it mean?
- When is it trustworthy?
- How can it be formalised?
- What assumptions are required?
- How can they be justified?

A Way Ahead?

A. Use contextual understanding of problem to justify causal input assumptions
   - randomization
   - Mendelian randomization
   - natural experiments
B. Perform various real experiments
   - Hunt for ECI properties
     - e.g., $X \perp\!\!\!\perp F_Y \mid (W, F_Z)$
   - Apply “causal discovery” to construct augmented DAG

A Parting Caution

- We have powerful statistical methods for framing and attacking causal problems
- To apply them, we need to make strong assumptions (e.g., relating different regimes)
- It is important to consider and justify these in any application

"No Causes In, No Causes Out"

Thank you!

Further Reading

- Dawid, A. P. (2002). Influence diagrams for causal modelling and inference. *Intern. Statist. Rev.* **70**, 161– 189. Corrigenda, ibid., 437.
- Dawid, A. P. (2003). Causal inference using influence diagrams: The problem of partial compliance (with Discussion). In *Highly Structured Stochastic Systems (*P. J. Green, N. L. Hjort and S. Richardson, Eds.). Oxford University Press, 45 – 81.
- Dawid, A. P. (2004). Probability, causality and the empirical world: A Bayes–de Finetti–Popper–Borel synthesis. *Statistical Science* **19**, 44 – 57.
- Dawid, A. P. (2007). Fundamentals Of Statistical Causality. Research Report 279, Department of Statistical Science, University College London.
- Dawid, A. P. (2010). Beware of the DAG! In Proceedings of the NIPS 2008 Workshop on Causality (I. Guyon, D. Janzing and B. Schölkopf,, Eds.). Journal of Machine Learning Research Workshop and Conference Proceedings 6, 59–86. http://tinyurl.com/33va7tm
- Dawid, A. P. (2010). Seeing and doing: The Pearlian synthesis. In Heuristics, Probability and Causality: A Tribute to Judea Pearl (R. Dechter, H. Geffner and J. Y. Halpern, Eds.). London: College Publications, 309–325.