Causal Inference from a Logical Point of View

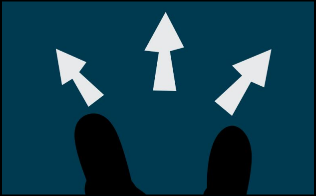
IAS Workshop on Logic & AI

Thomas Icard

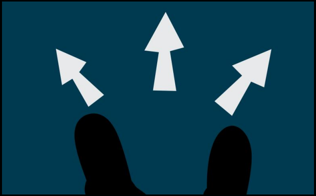
July 17, 2024

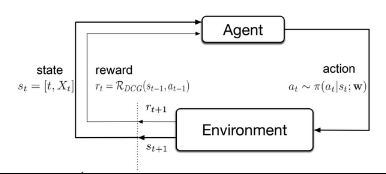
Causality as a bridge between Logic & AI

Causality as a bridge between Logic & AI

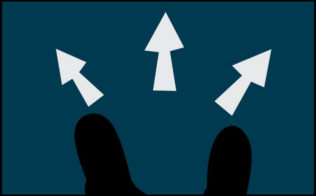


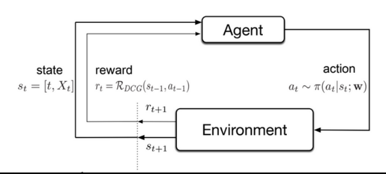
Causality as a bridge between Logic & AI



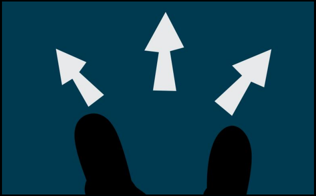


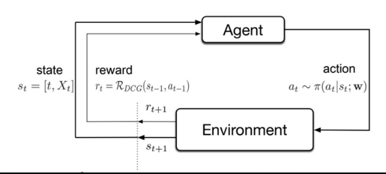
Causality as a bridge between Logic & AI





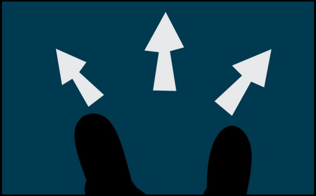
Causality as a bridge between Logic & AI

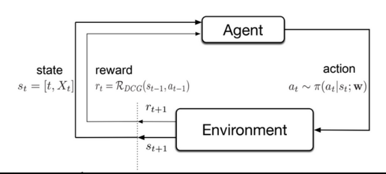




Hard-wired or learned?

Causality as a bridge between Logic & AI





Hard-wired or learned?

Cf., e.g., Richens & Everitt, *ICLR* 2024

Typical causal inference questions

- What is the impact of mask mandates on covid-19 cases?
- Will raising the minimum wage increase, decrease, or have no effect on unemployment?
- What could this study performed on rats tell us about the human immune system?
- How will presenting an advertisement here affect a consumer's purchasing behavior?
- Would the patient have survived had they not been given the treatment?

$$\text{Assumptions} + \text{Data} \overset{?}{\models} \text{Causal Conclusion}$$

$$\text{Assumptions} \; + \; \text{Data} \; \overset{?}{\models} \; \text{Causal Conclusion}$$

$$\left\{ \begin{array}{c} \text{functional} \\ \text{graphical} \\ \text{noise} \\ \vdots \end{array} \right\} \left\{ \begin{array}{c} \text{observation} \\ \text{experiment} \\ \text{quasi-exp.} \\ \vdots \end{array} \right\} \left\{ \begin{array}{c} \text{bounds on causal effects} \\ \text{causal direction} \\ \text{counterfactual probabilities} \\ \vdots \end{array} \right\}$$

$$\text{Assumptions} + \text{Data} \overset{?}{\models} \text{Causal Conclusion}$$

$$\left\{\begin{array}{c}\text{functional}\\ \text{graphical}\\ \text{noise}\\ \vdots\end{array}\right\} \left\{\begin{array}{c}\text{observation}\\ \text{experiment}\\ \text{quasi-exp.}\\ \vdots\end{array}\right\} \left\{\begin{array}{c}\text{bounds on causal effects}\\ \text{causal direction}\\ \text{counterfactual probabilities}\\ \vdots\end{array}\right\}$$

Logical angle: make syntax and semantics explicit.

Semantics: Structural Causal Models

$$\mathcal{M} \;=\; (\mathbf{V}, \mathbf{U}, \mathcal{F}, P)$$

Semantics: Structural Causal Models

$$\mathcal{M} \ = \ (\mathbf{V}, \mathbf{U}, \mathcal{F}, P)$$

- $\mathbf{V} = \{X, Y, Z, \dots\}$ endogenous variables
- $\mathbf{U} = \{U_1, U_2, U_3, \dots\}$ exogenous variables

Semantics: Structural Causal Models

$$\mathcal{M} \;=\; (\mathbf{V}, \mathbf{U}, \mathcal{F}, P)$$

- $\mathbf{V} = \{X, Y, Z, \dots\}$ endogenous variables
- $\mathbf{U} = \{U_1, U_2, U_3, \dots\}$ exogenous variables
- Each $X \in \mathbf{V}$ has *structural function* $f_X \in \mathcal{F}$:

$$f_X : \mathsf{Val}(\mathbf{V} \cup \mathbf{U}) \to \mathsf{Val}(X)$$

Semantics: Structural Causal Models

$$\mathcal{M} \;=\; (\mathbf{V}, \mathbf{U}, \mathcal{F}, P)$$

- $\mathbf{V} = \{X, Y, Z, \ldots\}$ endogenous variables
- $\mathbf{U} = \{U_1, U_2, U_3, \ldots\}$ exogenous variables
- Each $X \in \mathbf{V}$ has *structural function* $f_X \in \mathcal{F}$:

$$f_X : \mathrm{Val}(\mathbf{V} \cup \mathbf{U}) \to \mathrm{Val}(X)$$

- $P(\mathbf{U})$ is a probability distribution.

Front-door graph (Pearl 1995)

$$\begin{tikzcd} & & U \arrow[dll, dotted] \arrow[drr, dotted] & & \\ X \arrow[rr] & & Z \arrow[rr] & & Y \end{tikzcd} \quad + \quad P(X, Y, Z) \models E(Y_x - Y_{x'})$$

$$\Downarrow$$

$$P(x_z) = P(x)$$
$$P(z_x) = P(z|x)$$
$$P(y_z|x_z) = P(y|x, z)$$
$$P(y_x|z_x) = P(y_{x,z}) = P(y_z)$$

Observational data

Causal effect

Local ATE (Angrist & Imbens 1994)

Exclusion restriction $P(y_{x,z}, y'_{x,z'}) = 0$ Monotonicity $P(x'_z, x_{z'}) = 0$	$+$	(quasi-)exp. data $P(X_z), P(Y_z)$	$\models$	LATE $E(Y_x - Y_{x'} \mid x_z, x'_{z'})$

Logical Syntax

Logical Syntax

$$\varphi \;::=\; \mathbf{t} \succsim \mathbf{t} \;\mid\; \neg\varphi \;\mid\; \varphi \wedge \varphi$$

Logical Syntax

$$\varphi \;::=\; \mathbf{t} \succsim \mathbf{t} \;\mid\; \neg\varphi \;\mid\; \varphi \wedge \varphi$$

$$\mathbf{t} \;::=\; \mathbf{P}(\beta) \;\mid\; \dots$$

Logical Syntax

$$\varphi \;::=\; \mathbf{t} \succsim \mathbf{t} \;\mid\; \neg\varphi \;\mid\; \varphi \wedge \varphi$$

$$\mathbf{t} \;::=\; \mathbf{P}(\beta) \;\mid\; \dots$$

How much arithmetic do we admit?

Which β do we allow in $\mathbf{P}(\beta)$?

Complexity

Complexity

$$
\begin{tikzcd}
\mathsf{SAT}^{\mathrm{cond}} \arrow[rr] & & \mathsf{SAT}^{\mathrm{poly}} \\
& \exists\mathbb{R} & \\
& \mathsf{NP} & \\
\mathsf{SAT}^{\mathrm{comp}} \arrow[rr] \arrow[uuu] & & \mathsf{SAT}^{\mathrm{lin}} \arrow[uuu]
\end{tikzcd}
$$

(Mossé, Ibeling, & I., 2024)

Logical Syntax

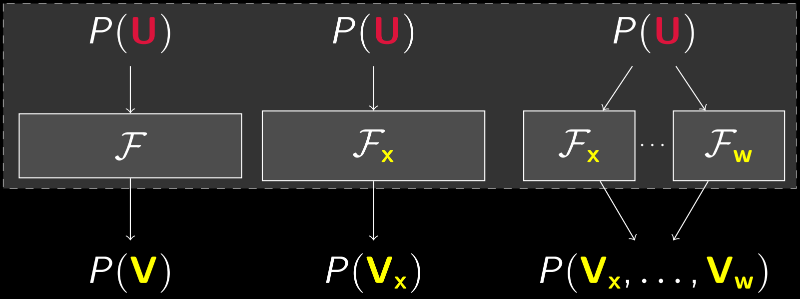
$$\varphi \;::=\; \mathbf{t} \succsim \mathbf{t} \;\mid\; \neg\varphi \;\mid\; \varphi \wedge \varphi$$

$$\mathbf{t} \;::=\; \underline{\mathbf{P}(\beta)} \;\mid\; \dots$$

How much arithmetic do we admit?

Which β do we allow in $\mathbf{P}(\beta)$?

Observational Interventional Counterfactual



(cf. Pearl & MacKenzie 2018; Bareinboim, Correa, Ibeling & I. 2020)

Counterfactual questions

Counterfactual questions

"For a patient who survived after treatment, how likely would they have survived if withheld treatment?"

Counterfactual questions

"For a patient who survived after treatment, how likely would they have survived if withheld treatment?"

$\approx$

"Did the patient survive because of the treatment?"

Counterfactual questions

"For a patient who survived after treatment, how likely would they have survived if withheld treatment?"

$$\approx$$

"Did the patient survive because of the treatment?"

$$P(Y_{X=0} = 0 \mid Y_{X=1} = 1)$$

Counterfactual questions

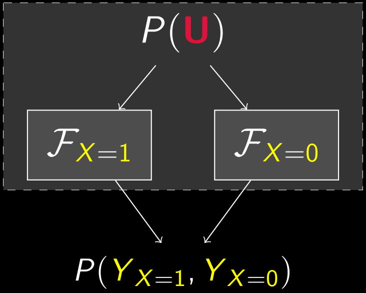
"For a patient who survived after treatment, how likely would they have survived if withheld treatment?"

$$\approx$$

"Did the patient survive because of the treatment?"

$$P(Y_{X=0} = 0 \mid Y_{X=1} = 1) \quad = \quad \frac{P(Y_{X=0} = 0, Y_{X=1} = 1)}{P(Y_{X=1} = 1)}$$

Counterfactual Distribution



$$\varphi ::= \mathbf{t} \succsim \mathbf{t} \mid \neg\varphi \mid \varphi \wedge \varphi$$

$$\mathbf{t} ::= \underline{\mathbf{P}(\beta)} \mid \mathbf{t} + \mathbf{t} \mid \mathbf{t} \cdot \mathbf{t}$$

$$\varphi \ ::= \ \mathbf{t} \succsim \mathbf{t} \ \mid \ \neg\varphi \ \mid \ \varphi \wedge \varphi$$

$$\mathbf{t} \ ::= \ \underline{\mathbf{P}(\beta)} \ \mid \ \mathbf{t} + \mathbf{t} \ \mid \ \mathbf{t} \cdot \mathbf{t}$$

$\mathcal{L}_1$: β propositional, e.g., $Y = 1 \wedge X = 1$.

$$\varphi \ ::= \ \mathbf{t} \succsim \mathbf{t} \ \mid \ \neg\varphi \ \mid \ \varphi \wedge \varphi$$

$$\mathbf{t} \ ::= \ \underline{\mathbf{P}(\beta)} \ \mid \ \mathbf{t} + \mathbf{t} \ \mid \ \mathbf{t} \cdot \mathbf{t}$$

$\mathcal{L}_1$: β propositional, e.g., $Y = 1 \wedge X = 1$.

$\mathcal{L}_2$: β a potential outcome, e.g., $Y_{X=1} = 1$.

$$\varphi \;::=\; \mathbf{t} \succsim \mathbf{t} \;\mid\; \neg\varphi \;\mid\; \varphi \wedge \varphi$$

$$\mathbf{t} \;::=\; \underline{\mathbf{P}(\beta)} \;\mid\; \mathbf{t} + \mathbf{t} \;\mid\; \mathbf{t} \cdot \mathbf{t}$$

$\mathcal{L}_1$: β propositional, e.g., $Y = 1 \wedge X = 1$.

$\mathcal{L}_2$: β a potential outcome, e.g., $Y_{X=1} = 1$.

$\mathcal{L}_3$: β a Boolean combination of potential outcomes, e.g., $Y_{X=0} = 0 \wedge Y_{X=1} = 1$.

Some logical observations

Some logical observations

- All three are nicely and simply axiomatizable (Ibeling et al., 2020, 2024).

Some logical observations

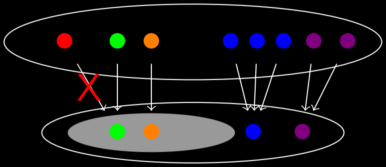
- All three are nicely and simply axiomatizable (Ibeling et al., 2020, 2024).
- Complexity depends only on arithmetic, not on causal sophistication (Mossé et al. 2024).

Some logical observations

- All three are nicely and simply axiomatizable (Ibeling et al., 2020, 2024).
- Complexity depends only on arithmetic, not on causal sophistication (Mossé et al. 2024).
- Formalizing derivations pinpoints hidden assumptions (Ibeling & I., 2023).

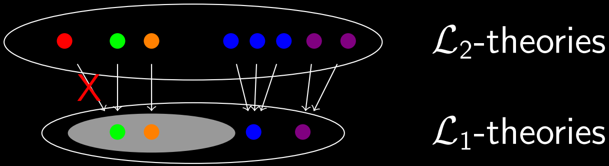
Some logical observations

- All three are nicely and simply axiomatizable (Ibeling et al., 2020, 2024).
- Complexity depends only on arithmetic, not on causal sophistication (Mossé et al. 2024).
- Formalizing derivations pinpoints hidden assumptions (Ibeling & I., 2023).
- Some current questions: marginalization/summation, etc.



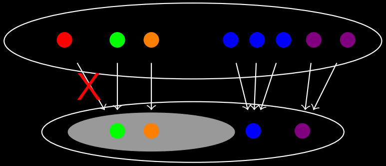
$\mathcal{L}_2$-theories

$\mathcal{L}_1$-theories



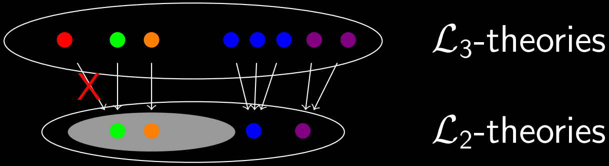
Theorem (Bareinboim, Correa, Ibeling, & I. 2022)

$\mathcal{L}_2$ *never* collapses into $\mathcal{L}_1$. The collapse set is empty.



$\mathcal{L}_3$-theories

$\mathcal{L}_2$-theories



Theorem (Ibeling & I. 2021)

The collapse set is meagre (in fact, nowhere dense) in the space of $\mathcal{L}_2$-theories with the weak topology.

Explanation

Explanation

- simplicity

Explanation

- simplicity
- generality

Explanation

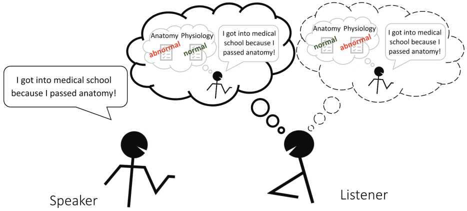
- simplicity
- generality
- unification

Explanation

- simplicity
- generality
- unification
- level of abstraction

Explanation

- simplicity
- generality
- unification
- level of abstraction



(See Kirfel et al. 2022; Harding et al. *in prep*)

Conclusion

Conclusion

- Causality is a fundamental ingredient for many AI tasks, connected with action, generalization & transfer, explanation, etc.

Conclusion

- Causality is a fundamental ingredient for many AI tasks, connected with action, generalization & transfer, explanation, etc.
- The logic of causality brings a rich set of mathematical and conceptual tools, raising many new questions for logic itself.

Conclusion

- Causality is a fundamental ingredient for many AI tasks, connected with action, generalization & transfer, explanation, etc.
- The logic of causality brings a rich set of mathematical and conceptual tools, raising many new questions for logic itself.
- Prediction: this will remain a fruitful point of contact between logic & AI.

Thanks for your attention!

Example

$$X := U_1 \qquad\qquad U_1 \sim \text{Bernoulli}(0.5)$$
$$Y := U_2 \qquad\qquad U_2 \sim \text{Bernoulli}(0.5)$$

Example

$$\begin{aligned} X &:= U_1 \qquad & U_1 &\sim \text{Bernoulli}(0.5) \\ Y &:= U_2 \qquad & U_2 &\sim \text{Bernoulli}(0.5) \end{aligned}$$

$$P(Y_{X=1} = 1) \;=\; 1/2$$

Example

$$\begin{aligned} X &:= U_1 & \qquad U_1 &\sim \text{Bernoulli}(0.5) \\ Y &:= U_2 & \qquad U_2 &\sim \text{Bernoulli}(0.5) \end{aligned}$$

$$\begin{aligned} P(Y_{X=1} = 1) &= 1/2 \\ P(Y_{X=1} = 1 \mid Y = 0, X = 0) &= 0 \end{aligned}$$

Example

$$\begin{aligned} X &:= U_1 & \qquad U_1 &\sim \text{Bernoulli}(0.5) \\ Y &:= (X \leftrightarrow U_2) & \qquad U_2 &\sim \text{Bernoulli}(0.5) \end{aligned}$$

$$P(Y_{X=1} = 1) \;=\; 1/2$$

Example

$$
\begin{aligned}
X &:= U_1 & \qquad U_1 &\sim \text{Bernoulli}(0.5) \\
Y &:= (X \leftrightarrow U_2) & \qquad U_2 &\sim \text{Bernoulli}(0.5)
\end{aligned}
$$

$$P(Y_{X=1} = 1) \;=\; 1/2$$

$$P(Y_{X=1} = 1 \mid Y = 0, X = 0)$$

Example

$$\begin{aligned} X &:= U_1 \\ Y &:= (X \leftrightarrow U_2) \end{aligned} \qquad \begin{aligned} U_1 &\sim \text{Bernoulli}(0.5) \\ U_2 &\sim \text{Bernoulli}(0.5) \end{aligned}$$

$$\begin{aligned} P(Y_{X=1} = 1) &= 1/2 \\ P(Y_{X=1} = 1 \mid Y = 0, X = 0) &= 1 \end{aligned}$$

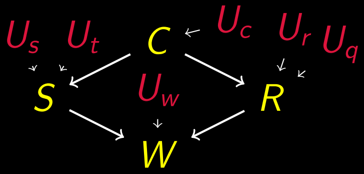
Effect of Treatment on Treated (ETT)

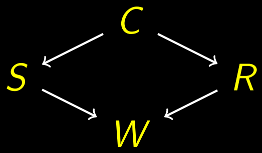
$$\begin{aligned} \mathbf{P}(y_x) &= \mathbf{P}(y_x \wedge x') + \mathbf{P}(y_x \wedge x) \\ &= \mathbf{P}(y_x \wedge x') + \mathbf{P}(y \wedge x) \\ &= \mathbf{P}(y_x | x')\mathbf{P}(x') + \mathbf{P}(y \wedge x) \end{aligned}$$

Hence:

$$\mathbf{P}(y_x | x') = \frac{\mathbf{P}(y_x) - \mathbf{P}(y \wedge x)}{\mathbf{P}(x')}$$

$$
\begin{aligned}
C &:= U_c & U_c &\sim \mathrm{Bern}(0.5)\\
S &:= (C \wedge U_s) \vee & U_s &\sim \mathrm{Bern}(0.1)\\
&\quad\ (\neg C \wedge U_t) & U_t &\sim \mathrm{Bern}(0.5)\\
R &:= (C \wedge U_r) \vee & U_r &\sim \mathrm{Bern}(0.8)\\
&\quad\ (\neg C \wedge U_q) & U_q &\sim \mathrm{Bern}(0.2)\\
W &:= (S \vee R) \wedge U_w & U_w &\sim \mathrm{Bern}(0.95)
\end{aligned}
$$




C
S
R
W

$$\begin{tikzcd}
 & C \arrow[dl] \arrow[dr] & \\
S \arrow[dr] & & R \arrow[dl] \\
 & W &
\end{tikzcd}$$

	1	0
$P(C)$	0.5	0.5

	1	0
$P(S \mid C = 1)$	0.1	0.9
$P(S \mid C = 0)$	0.5	0.5

	1	0
$P(R \mid C = 1)$	0.8	0.2
$P(R \mid C = 0)$	0.2	0.8

	1	0
$P(W \mid S + R \neq 0)$	0.95	0.05
$P(W \mid S = R = 0))$	0.0	1.0

$$
\begin{aligned}
C &:= U_c \\
S &:= (C \wedge U_s) \vee \\
&\quad\ (\neg C \wedge U_t) \\
R &:= (C \wedge U_r) \vee \\
&\quad\ (\neg C \wedge U_q) \\
W &:= (S \vee R) \wedge U_w
\end{aligned}
\qquad
\begin{aligned}
U_c &\sim \text{Bern}(0.5) \\
U_s &\sim \text{Bern}(0.1) \\
U_t &\sim \text{Bern}(0.5) \\
U_r &\sim \text{Bern}(0.8) \\
U_q &\sim \text{Bern}(0.2) \\
U_w &\sim \text{Bern}(0.95)
\end{aligned}
$$

$$P(S_{C=0} \mid C = S = 1)$$

$$
\begin{aligned}
C &:= U_c \\
S &:= (C \wedge U_s) \vee \\
&\quad\;\; (\neg C \wedge U_t) \\
R &:= (C \wedge U_r) \vee \\
&\quad\;\; (\neg C \wedge U_q) \\
W &:= (S \vee R) \wedge U_w
\end{aligned}
\qquad
\begin{aligned}
U_c &\sim \mathrm{Bern}(0.5) \\
U_s &\sim \mathrm{Bern}(0.1) \\
U_t &\sim \mathrm{Bern}(0.5) \\
U_r &\sim \mathrm{Bern}(0.8) \\
U_q &\sim \mathrm{Bern}(0.2) \\
U_w &\sim \mathrm{Bern}(0.95)
\end{aligned}
$$

$$P(S_{C=0} \mid C = S = 1) \;=\; P(S_{C=0})$$

$$
\begin{aligned}
C &:= U_c & U_c &\sim \text{Bern}(0.5)\\
S &:= (C \wedge U_s = 1) \vee & U_s &\sim \text{Unif}(1, 10)\\
&\quad\ (\neg C \wedge U_s \leq 5) & &\\
R &:= (C \wedge U_r) \vee & U_r &\sim \text{Bern}(0.8)\\
&\quad\ (\neg C \wedge U_q) & U_q &\sim \text{Bern}(0.2)\\
W &:= (S \vee R) \wedge U_w & U_w &\sim \text{Bern}(0.95)
\end{aligned}
$$

$$
\begin{aligned}
C &:= U_c & U_c &\sim \mathrm{Bern}(0.5)\\
S &:= (C \wedge U_s = 1) \vee & U_s &\sim \mathrm{Unif}(1, 10)\\
&\quad\ (\neg C \wedge U_s \leq 5) & &\\
R &:= (C \wedge U_r) \vee & U_r &\sim \mathrm{Bern}(0.8)\\
&\quad\ (\neg C \wedge U_q) & U_q &\sim \mathrm{Bern}(0.2)\\
W &:= (S \vee R) \wedge U_w & U_w &\sim \mathrm{Bern}(0.95)
\end{aligned}
$$

$$P(S_{C=0} \mid C = S = 1)$$

$$
\begin{aligned}
C &:= U_c & U_c &\sim \mathrm{Bern}(0.5)\\
S &:= (C \wedge U_s = 1) \vee & U_s &\sim \mathrm{Unif}(1, 10)\\
&\quad\ (\neg C \wedge U_s \leq 5) & &\\
R &:= (C \wedge U_r) \vee & U_r &\sim \mathrm{Bern}(0.8)\\
&\quad\ (\neg C \wedge U_q) & U_q &\sim \mathrm{Bern}(0.2)\\
W &:= (S \vee R) \wedge U_w & U_w &\sim \mathrm{Bern}(0.95)
\end{aligned}
$$

$$
\begin{aligned}
P(S_{C=0} \mid C = S = 1) &= 1\\
&\neq P(S_{C=0})
\end{aligned}
$$

$\mathcal{L}_1$ – “pure” probability logic

$\mathcal{L}_1$ – "pure" probability logic

- Transitivity, comparability of $\succsim$
- Boolean reasoning
- $\mathbf{P}(\alpha) \succsim \mathbf{P}(\beta)$ whenever $\models \beta \to \alpha$
- $\neg 0 \succsim 1$
- $\mathbf{P}(\alpha) \approx \mathbf{P}(\alpha \wedge \beta) + \mathbf{P}(\alpha \wedge \neg\beta)$

$\mathcal{L}_1$ –

$$a + (b + c) \approx (a + b) + c$$
$$a + b \approx b + a$$
$$a + 0 \approx a$$
$$(a + e \succsim c + f \wedge b + f \succsim d + e) \to a + b \succsim c + d$$
$$(a + b \succsim c + d \wedge d \succsim b) \to a \succsim c$$

$\mathcal{L}_1$ –

$$\mathsf{a} + (\mathsf{b} + \mathsf{c}) \approx (\mathsf{a} + \mathsf{b}) + \mathsf{c}$$
$$\mathsf{a} + \mathsf{b} \approx \mathsf{b} + \mathsf{a}$$
$$\mathsf{a} + 0 \approx \mathsf{a}$$
$$(\mathsf{a} + \mathsf{e} \succsim \mathsf{c} + \mathsf{f} \wedge \mathsf{b} + \mathsf{f} \succsim \mathsf{d} + \mathsf{e}) \to \mathsf{a} + \mathsf{b} \succsim \mathsf{c} + \mathsf{d}$$
$$(\mathsf{a} + \mathsf{b} \succsim \mathsf{c} + \mathsf{d} \wedge \mathsf{d} \succsim \mathsf{b}) \to \mathsf{a} \succsim \mathsf{c}$$

$$\mathsf{a} \cdot (\mathsf{b} \cdot \mathsf{c}) \approx (\mathsf{a} \cdot \mathsf{b}) \cdot \mathsf{c}$$
$$\mathsf{a} \cdot \mathsf{b} \approx \mathsf{b} \cdot \mathsf{a}$$
$$\mathsf{a} \cdot 0 \approx 0$$
$$\mathsf{a} \cdot 1 \approx \mathsf{a}$$
$$\mathsf{c} \succ 0 \to (\mathsf{a} \cdot \mathsf{c} \succsim \mathsf{b} \cdot \mathsf{c} \leftrightarrow \mathsf{a} \succsim \mathsf{b})$$
$$\mathsf{a} \cdot (\mathsf{b} + \mathsf{c}) \approx \mathsf{a} \cdot \mathsf{b} + \mathsf{a} \cdot \mathsf{c}$$
$$\mathsf{a} \succsim \mathsf{b} \wedge \mathsf{c} \succsim \mathsf{d} \to \mathsf{a} \cdot \mathsf{c} + \mathsf{b} \cdot \mathsf{d} \succsim \mathsf{a} \cdot \mathsf{d} + \mathsf{b} \cdot \mathsf{c}$$

$\mathcal{L}_3$ – probabilistic logic of counterfactuals

$\mathcal{L}_3$ – probabilistic logic of counterfactuals

$\mathbf{P}(\alpha) \gtrsim \mathbf{P}(\beta)$ whenever $\models_{\text{cond}} \beta \rightarrow \alpha$

$\mathcal{L}_3$ – probabilistic logic of counterfactuals

$$\mathbf{P}(\alpha) \gtrsim \mathbf{P}(\beta) \text{ whenever } \models_{\text{cond}} \beta \rightarrow \alpha$$

$$\bigwedge_{i=1}^{n-1} \mathbf{P}\big([\alpha_i \wedge X_i] I_{i+1} \wedge [\alpha_i \wedge \neg X_i] \neg I_{i+1}\big) \succ 0$$

$$\rightarrow \; \mathbf{P}\big([\alpha_k \wedge X_k] I_1 \wedge [\alpha_k \wedge \neg X_k] \neg I_1\big) \approx 0$$

(Ibeling & I., 2020)