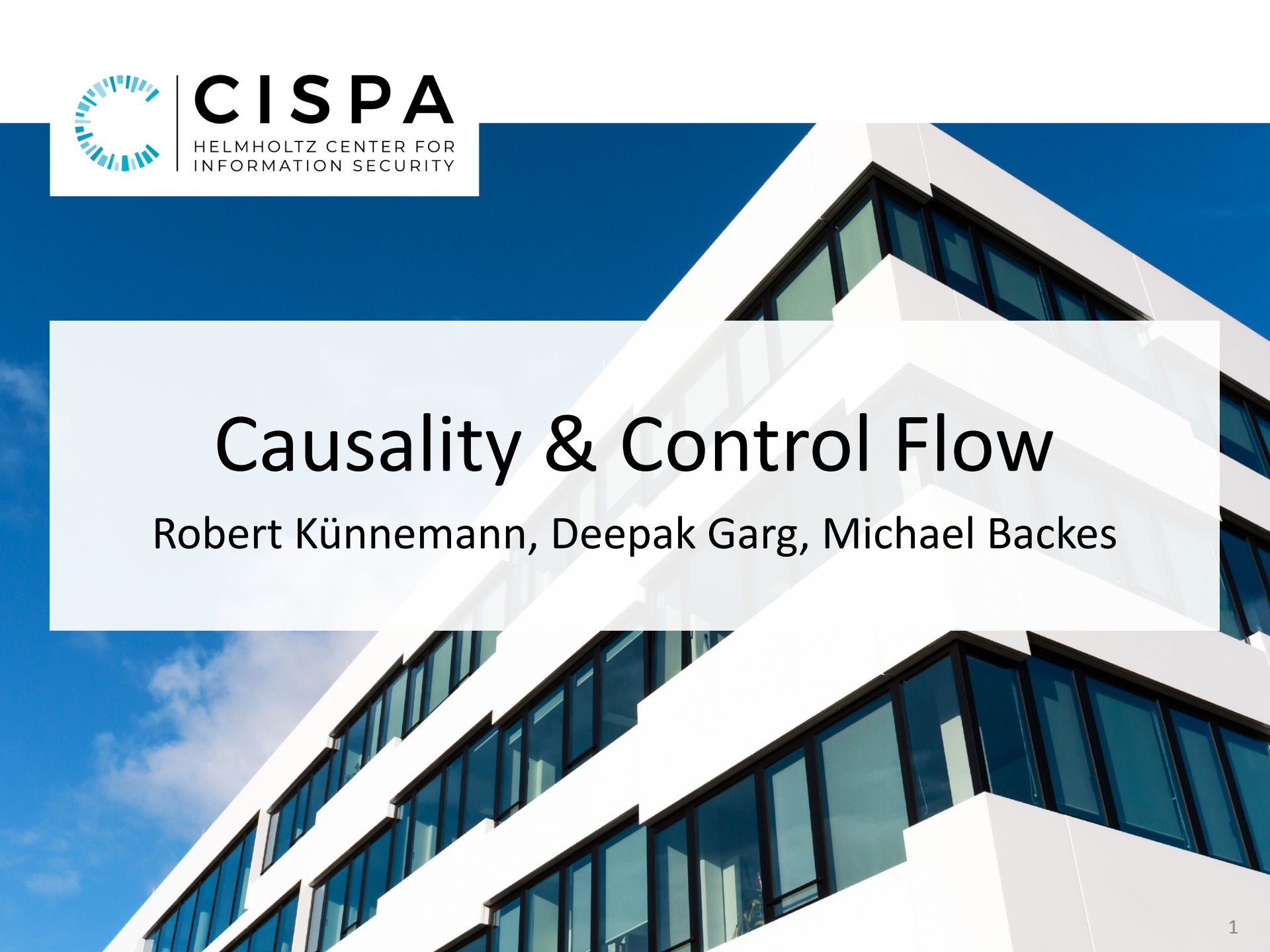
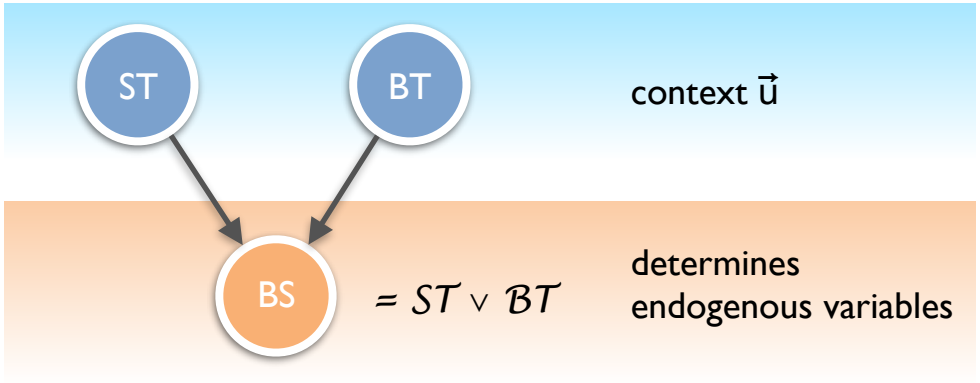


Causality & Control Flow

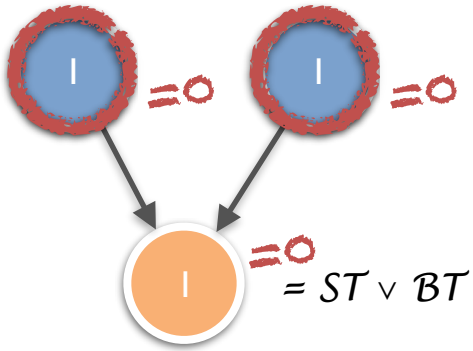
Robert Künnemann, Deepak Garg, Michael Backes

Pearl's structured equations model

- Random variables: exogenous (outside model), endogenous



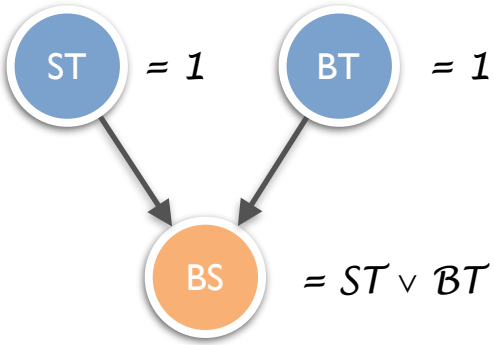
- Intervention: express what-if test

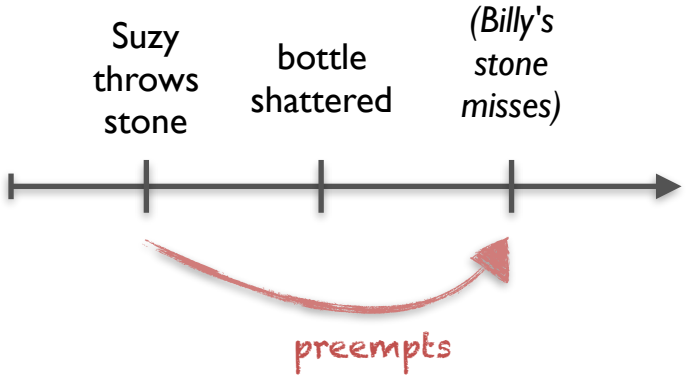


- $\vec{X}=\vec{x}$ what-if cause of Y=y in context $\vec{u}$ iff
  - $\vec{u} \models \vec{X}=\vec{x} \wedge Y=y$
  - $\vec{u} \models [\vec{X} \leftarrow \vec{x}'] \; Y \neq y$
  - $\vec{X}$ minimal

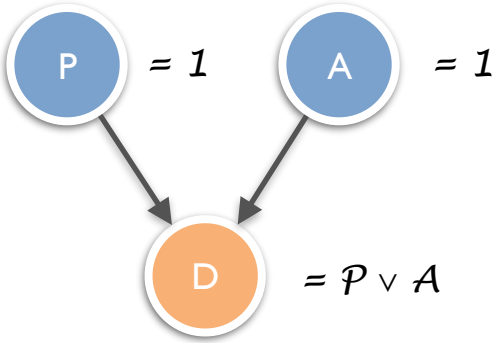
Preemption

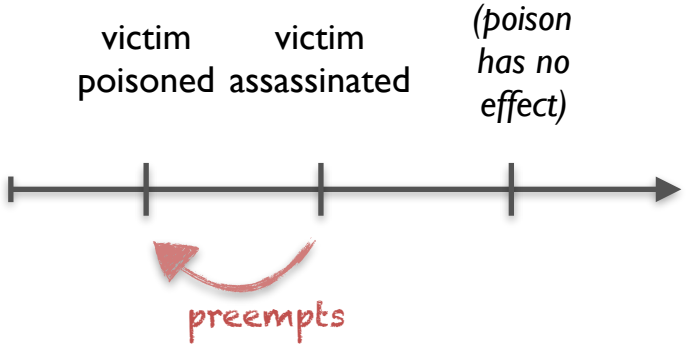
late preemption





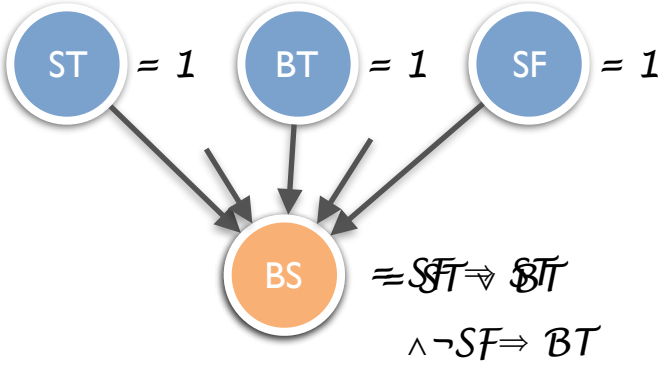
early preemption





How to model control-flow

- Where should it be represented
  - equations: … no
  - context: .. no
  - control-flow variables

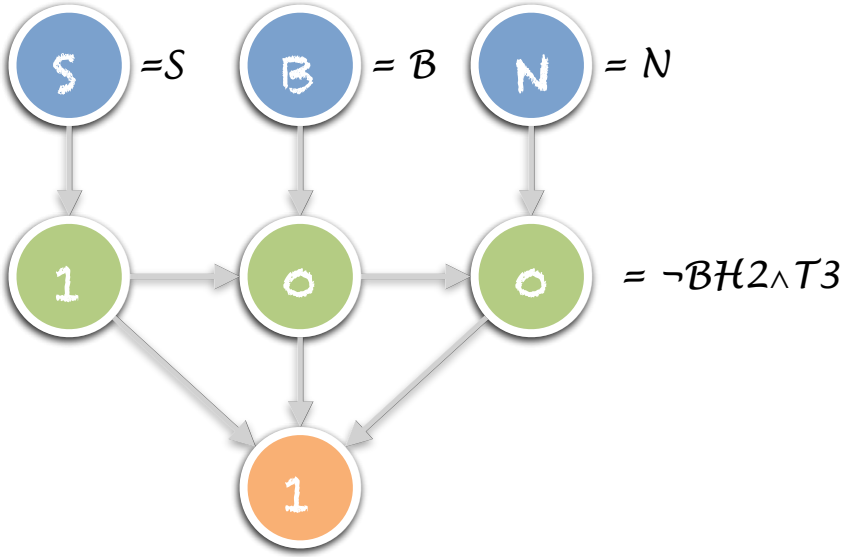


- Intervention only on data-flow
  - causes can only be about data-flow
  - events can change, but temporal relations remain fixed

- Key-question: how does counterfactual control-flow relate to actual control-flow?

Our contribution: How to model control-flow

- Where should it be represented
  - equations: … no
  - context: .. no
  - control-flow variables

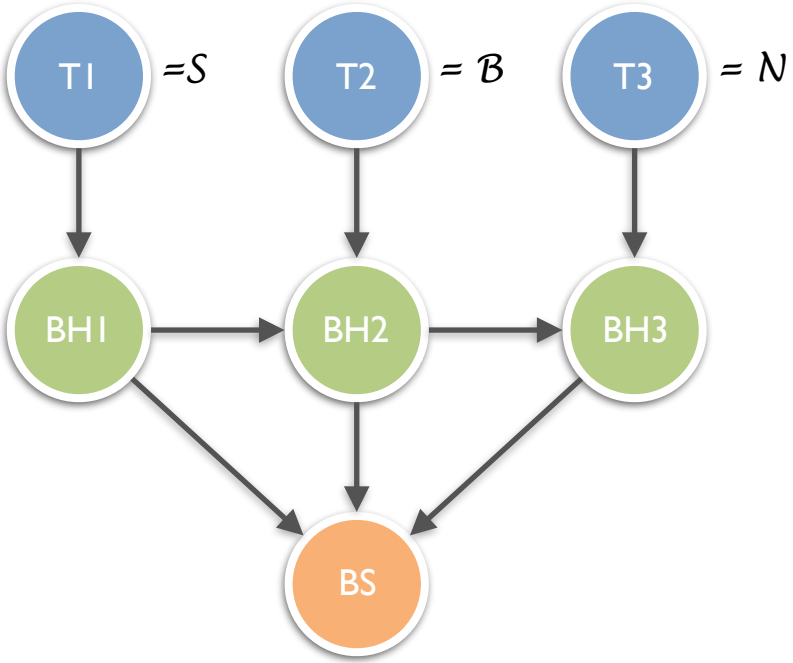


- Intervention only on data-flow
  - causes can only be about data-flow
  - events can change, but temporal relations remain fixed

- Key-question: how does counterfactual control-flow relate to actual control-flow?

How to model control-flow

- Where should it be represented

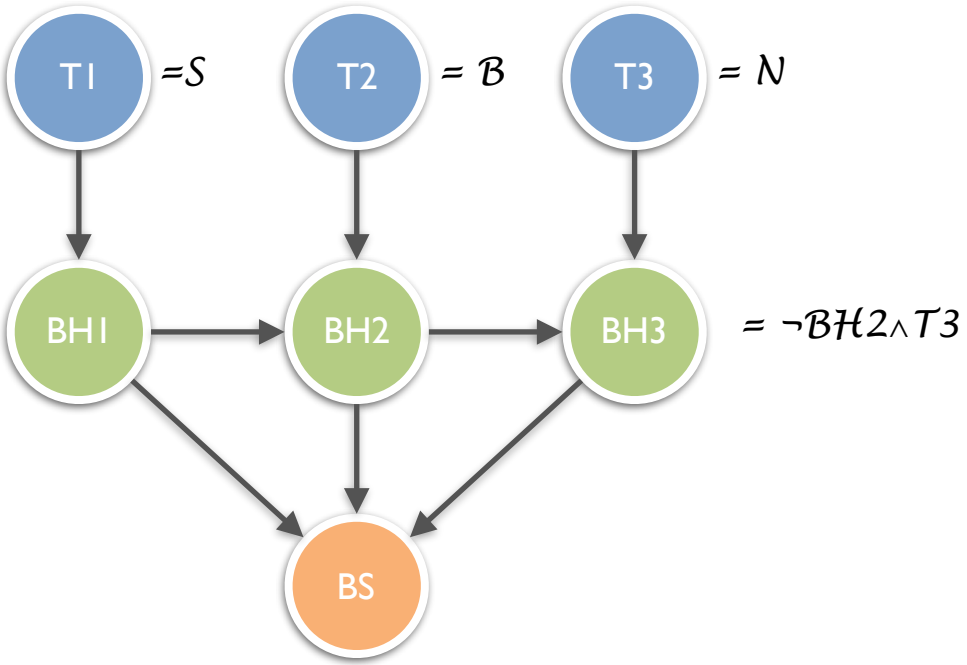



```
if T1=S or T1=B
then
    BS=1
else if T2=S or T2=B
    then
        BS=1
    else
        if T3=S or T3=B
        then
            BS=1
        else
            BS=0
```

- Key-question: how does counterfactual control-flow relate to actual control-flow?

How to model control-flow

- Where should it be represented
  - equations: … no
  - context: .. no
  - control-flow variables

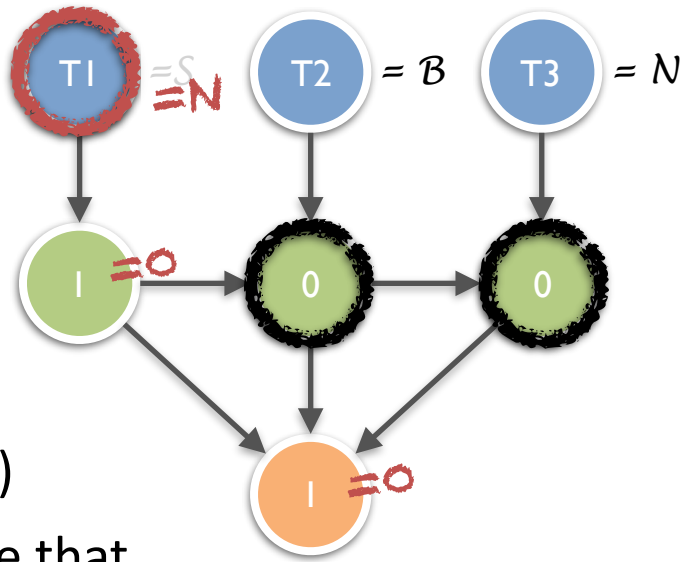


- Intervention only on data-flow
  - events can change,
    but temporal relations remain fixed
  - causes can only concern data-flow variable

- Key-question: how does counterfactual control-flow relate to actual control-flow?

Digression: Review Pearl/Halpern definition

- $\vec{X}=\vec{x}$ cause for Y=y iff
  - AC1, AC3 as before
  - AC2: $\exists \vec{W},\vec{w},\vec{x}.\ \vec{u} \models \vec{W}=\vec{w} \wedge [\vec{X}\leftarrow\vec{x}', \vec{W}\leftarrow\vec{w}]\ Y\neq y$



- generalisation of what-if cause ($\vec{W}$ empty)
- alternative reading: remove parts of cause that fix to actual values

- observation 1: contingencies concern control-flow in almost all examples
- observation 2: if they include control-flow, they can give unintuitive results
- consequence: restrict $\vec{W}$ to control-flow, $\vec{X}$ to data-flow variables

Back to key-question:

CFN2: For $\vec{C}$ the actual control flow (i.e., the ctl-flow variables set to 1), there is $\vec{x}'$ s.t. $\vec{u} \models [(V_{ctl} \setminus C) \leftarrow \vec{0}, \vec{X} \leftarrow \vec{x}'] Y \neq y$

Freedom in choosing ctl-flow

proposal	idea	drawback
Datta et.al. / Beckers	interventions need to preserve ctl-flow	coming about of event may be important
our proposal	may interrupt, but never stray from path	?
Lewis' what-if	arbitrary, but induced by data-flow	different course of events can mask preemption
Halpern'15	pick & choose	can pick inconsistent ctl-flow, e.g., Careful poisoning (Weslake)

Benchmarking

- Benchmarked with 34 Examples from Halpern'15 and Weslake'15 (no cherry-picking!)
- Available at https://github.com/rkunnema/causation-benchmark (Please extend!)

- **still not conclusive:**
  - flexibility in modelling
  - modelling principles, but very generic domain
- **how to get to conclusion?** translations, e.g., PDG to causal models

	Definition 4	Definition 7	Definition 6
Forest fire, Ex. 1 — disjunctive (overdet.) [9, p. 278] — disjunctive, ext. [11, Ex. 3.7]	(A,B,FF) (B,O), (S,O) (MD,L,C,FF), (MD,B,C,FF), (L,A,C,FF)	(A), (B), (FF) (B,S), (O) (MD), (L), (C), (FF)	(A,B,FF) (B,O), (S,O) (MD,L,FF)
Late preemption, Ex. 4 Early preemption [17, p. 526] — (ctl), Ex. 5[a]	$(T1,BS1,BS)$, $(T2,BS1,BS2,BS)$ (A,D_1,D_2), (B,P,D_2) (A,S_r,D), (B,P_r,S_r,PE_r,NS_r,D)	$(T1)$, $(BS1)$, (BS) (A), (D_1), (D_2) (A), (S_r), (D)	$(T1,BS)$ (A,D_2) (A,D)
Bogus prevention [15] — ad-hoc [13, p. 29] — (ctl), Ex. 6 — (ctl, reversed)[b]	(P,S), (A,S) (P,S), (A,S,PN) (P,S,NP_r), (A,S,P_r,NP_r,N_r) (P,D,NP_r), (A,D,NP_r,P_r,PN_r)	(P,A), (S) (P), (S) (P), (S), (NP_r) (P), (D), (NP_r)	(P,S), (A,S) (P,S) (P,S) (P,D)
Careful Poisoning [23, Ex. 11] — (ctl)[c]	(A,D), (P,D) (A,NA_r,P_r,D), (NA_r,A_r,P_r,P,D)	(A), (D) (A), (NA_r), (P_r), (D)	(A,D), (P,D) (D)
Train [12, Ex. 4] — [8] — (ctl)[d]	(F,RB,A), (LB,RB,A) (F,RT,A) (F,R_r,A), (L_r,R_r,A)	(F,LB), (RB), (A) (F), (RT), (A) (F), (R_r), (A)	(F,RB,A), (LB,RB,A) (F,A) (F,A)
Prisoner [18] Backup [23, Ex. 1] — (ctl) [23, Ex. 1] Command [23, Ex. 8]	(C,D) (T,V), (S,V) (T,V,T_r), (S,V,T_r,NT_r,S_r) (M,C)	(C), (D) (T), (V) (T), (V), (T_r) (M), (C)	(C,D) (T,V), (S,V) (T,V) (M,C)
Agreement, Ex. 7 BoS, Ex. 8	(A,B,R,O) (P_1,P_2,C,T)	(A), (B), (R), (O) (N), (P_1), (P_2), (C), (T)	(A,B,O) (P_1,P_2,T)
Switch [23, p. 16] Combination Lamp [23, p. 19] Shock [23, p. 17] Push A [23, p. 26] — (ctl)[e] Push B [23, p. 26] Fancy Lamp [23, p. 31]	$(S,L2,I)$, $(L1,L2,I)$ (B,C,L) $(B,C,C1)$ (P,B,H,D), (P,T,H,D) (P,T,P_r,H,D) (P,T,H,D) $(A,N3,L)$, $(B,N1,L)$	(S), $(L2)$, (I) (B), (C), (L) (A), (B), (C), $(C1)$ (P), (B,T), (H), (D) (P), (T), (P_r), (H), (D) (P), (T), (H), (D) (A,B), $(A,N1)$, $(B,N3)$, $(N1,N3)$, (L)	$(S,L2,I)$, $(L1,L2,I)$ (B,C,L) (B,C) (P,B,H,D), (P,T,H,D) (P,T,H,D) (P,T,H,D) $(A,N3,L)$, $(B,N1,L)$
Vote [11, Ex. 4.1] Ranch [11, Ex. 3.7] Vote 5:2[f]	(V_1,M,P), (V_2,M,P) (A_1,A_2,M_1,O) $(V1,V2,V3,V4,O)$, $(V1,V2,V3,V5,O)$, $(V1,V2,V4,V5,O)$, $(V1,V3,V4,V5,O)$, $(V2,V3,V4,V5,O)$	(V_1,V_2), (M), (P) (A_1), (A_2), (M_1), (O) $(V1,V2)$, $(V1,V3)$, $(V1,V4)$, $(V1,V5)$, $(V2,V3)$, $(V2,V4)$, $(V2,V5)$, $(V3,V4)$, $(V3,V5)$, $(V4,V5)$, (O)	(V_1,M,P), (V_2,M,P) (A_1,A_2) $(V1,V2,V3,V4)$, $(V1,V2,V3,V5)$, $(V1,V2,V4,V5)$, $(V1,V3,V4,V5)$, $(V2,V3,V4,V5)$
Pollution, $k = 80$ [11, Ex. 3.11] Pollution, $k = 50$ [11, Ex. 3.11] Pollution, $k = 120$ [11, Ex. 3.11]	(A,D) (A,D), (B,D) (A,B,D)	(A), (D) (A,B), (D) (A), (B), (D)	(A,D) (A,D), (B,D) (A,B,D)

Summary

- causality definitions struggle with preemption
- proposal: explicit modelling of control-flow
  - modelling principles
  - restriction on intervention
  - definition: control-flow preserving causation
- benefits:
  - model hard examples without secondary notions like normality
  - understand existing causality definitions better1

Thank you! Questions?