Efficient Computation of Actual HP Causality for Accountability

Amjad Ibrahim, Alexander Pretschner

Technische Universität München

fortiss research and technology transfer institute of the Free State of Bavaria

Bavarian Research Institute for Digital Transformation

Shonan, June 2019

Flavors of Causality

Spectrum-Based Fault Localization

Model-Based Diagnosis

Granger Causality

Halpern-Pearl Causality

The Three Layer Causal Hierarchy

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?

Figure 1: The Causal Hierarchy. Questions at level i can only be answered if information from level i or higher is available.

Flavors of Causality

Spectrum-Based Fault Localization

Model-Based Diagnosis

Granger Causality

Halpern-Pearl Causality

- **Definition**
- SAT-based computation
- ILP-based computation

The Three Layer Causal Hierarchy

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?

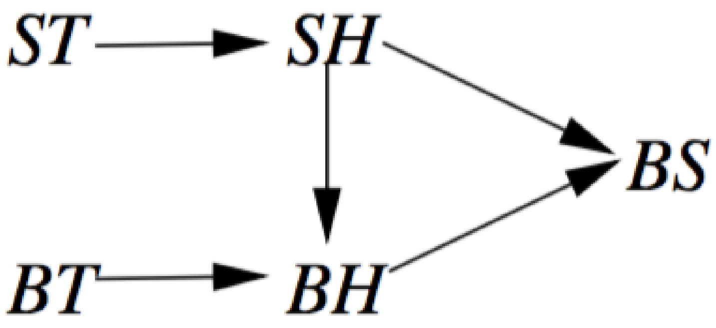
Figure 1: The Causal Hierarchy. Questions at level i can only be answered if information from level i or higher is available.

Actual causality based on Halpern and Pearl [HP]

- Remember counterfactual reasoning with but-for tests
- Causal models
  - Structural equations represent mechanisms of the world
  - Variables represent properties of the world
  - Interventions
  - Addresses the ‘problematic’ examples in literature
- Three versions: First (2001), Updated (2005), Modified (2015)
- We use it to explain failures, attacks and incidents
  - Attributing responsibility in malicious insiders attacks, CPS accidents

Causal Models

- Signature: S=(U,V,**R**)
  - U: Set of exogenous variables
  - V: Set of endogenous variables
  - **R**: Associates with each variable a set of possible values
- Causal Model: M=(S,**F**)
  - **F**: Associates a function F_X with each $X \in V$
    In words: „F_X tells us the value of X given the values of all other variables in $U \cup V$"
  - Visualization via Causal Networks



A Language for Causal Reasoning

- A sequence of variables $X_1,\ldots,X_n$ can be abbreviated using vector notation As $\vec{X}$. Analogously, $X_1 = x_1,\ldots, X_n = x_n$ is abbreviated $\vec{X} = \vec{x}$
- Variable Y can be set to value y writing $Y \leftarrow y$ (analogously $\vec{Y} \longleftarrow \vec{y}$ for vectors).
- Intervention on a model : setting the value of some variable *X to x* in a causal model *M = (S,F)* results in a new model $M_{X \leftarrow x}$ which is identical to *M* except that the equation for *X* in *F* is replaced with *X*=*x*
- A basic causal formula (over S) is of the form $[Y_1 \leftarrow y_1, Y_2 \leftarrow y_2, \ldots Y_n \leftarrow yn]\ \varphi$, where φ is a Boolean combination of primitive events, $[Y_1 \ldots Y_n]$(abbreviated $\vec{Y}$) are distinct variables in V
- A causal formula ψ can be evaluated to true or false in a causal model M given a context $\vec{u}$. We write $(M,\vec{u}) \vDash \psi$ if ψ evaluates to true in M
- $(M,\vec{u}) \vDash [\vec{Y} \longleftarrow \vec{y}](X \leftarrow x)$ implies that solving the equations in $M_{\vec{Y} \leftarrow \vec{y}}$

With context $\vec{u}$ yields the value x for variable X.

Modified HP Definition

$\vec{X} = \vec{x}$ is an actual cause of φ in $(M, \vec{u})$ if the following three conditions hold:

AC1. *(informal)* Both $\vec{X} = \vec{x}$ and φ need to actually happen.

Modified HP Definition

$\vec{X} = \vec{x}$ is an actual cause of φ in $(M, \vec{u})$if the following three conditions hold:

AC1. $(M, \vec{u}) \vDash (\vec{X} = \overrightarrow{x)}$ and $(M, \vec{u}) \vDash \varphi$

Modified HP Definition

$\vec{X} = \vec{x}$ is an actual cause of φ in $(M, \vec{u})$if the following three conditions hold:

AC1. $(M, \vec{u}) \vDash (\vec{X} = \overrightarrow{x)}$ and $(M, \vec{u}) \vDash \varphi$

AC3. $\vec{X}$ is minimal; no subset of $\vec{X}$ satisfies conditions AC1 and AC2.

Modified HP Definition

$\vec{X} = \vec{x}$ is an actual cause of φ in $(M, \vec{u})$if the following three conditions hold:

AC1. $(M, \vec{u}) \vDash (\vec{X} = \overrightarrow{x})$ and $(M, \vec{u}) \vDash \varphi$

AC2. *(informal)* Changing the original values of $\vec{X}$ to a different setting $\overrightarrow{x'}$ while keeping a possibly empty set of the remaining variables at their original value, φ does not occur anymore.

AC3. $\vec{X}$ is minimal; no subset of $\vec{X}$ satisfies conditions AC1 and AC2.

Modified HP Definition

$\vec{X} = \vec{x}$ is an actual cause of φ in $(M, \vec{u})$if the following three conditions hold:

AC1. $(M, \vec{u}) \vDash (\vec{X} = \overrightarrow{x})$ and $(M, \vec{u}) \vDash \varphi$

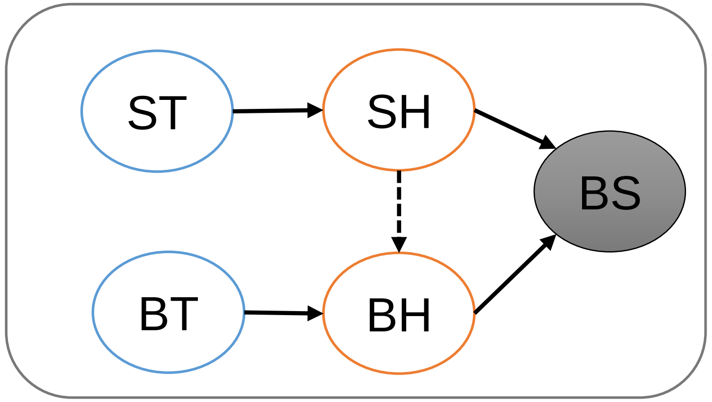
AC2. There is a set $\overrightarrow{W}$ of variables in V and a setting $\overrightarrow{x'}$of the variables in $\vec{X}$ such that if $(M, \vec{u}) \vDash (\overrightarrow{W} = \overrightarrow{w})$, then $(M, \vec{u}) \vDash \left[\vec{X} \leftarrow \overrightarrow{x'}, \overrightarrow{W} \leftarrow \overrightarrow{w}\right] \neg\varphi$

AC3. $\vec{X}$ is minimal; no subset of $\vec{X}$ satisfies conditions AC1 and AC2.

For binary models we have:

Lemma. If $\vec{X} = \vec{x}$ is an actual cause of φ in the sense of AC1-AC3, then every $\overrightarrow{x'}$ in the definition of AC2 always satisfies $\forall_i\ x'_i = \neg x_i$

Rock-Throwing Example

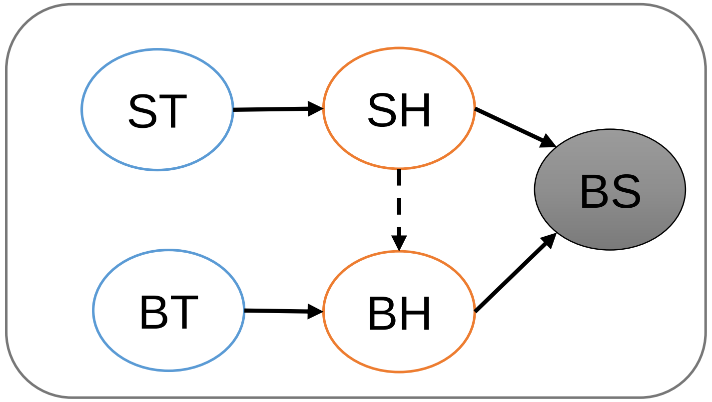


- ST/BT = Billy/Suzy throws
- SH = ST (Suzy hits)
- BH = BT ∧ ¬SH (Billy hits)
- BS = SH ∨ BH (Bottle shatters)

The real world:

- ST = BT = 1
- SH = ST = 1
- BH = BT ∧ ¬SH = 1 ∧ 0 = 0
- BS = SH ∨ BH = 1 ∨ 0 = 1

Rock-Throwing Example



- ST/BT = Billy/Suzy throws
- SH = ST
- BH = BT ∧ ¬SH
- BS = SH ∨ BH

AC2 (a^m): $(M, \vec{u}) \models [\vec{X} \leftarrow \vec{x}', \vec{W} \leftarrow \vec{w}]\neg\varphi$

Is ST a cause?
Set ST = 0 and $\vec{W} = \emptyset$
ST = 0; BT = 1
SH = ST = 0
BH = BT ∧ ¬SH = 1 ∧ 1 = 1
BS = SH ∨ BH = 0 ∨ 1 = 1
φ still occurs → AC2

Is ST a cause?
Set ST = 0 and $\vec{\boldsymbol{W}} = \{\mathbf{BH}\}$
ST = 0; BT = 1
SH = ST = 0
BH = 0
BS = SH ∨ BH = 0 ∨ 0 = 0
φ does not occur anymore → AC2

Practical Causal Inference

Problem:

- No comprehensive technical framework to model and benchmark causality inference
- Computational complexity of inferring actual causality is bad: worse than NP [11]; NP-complete for special cases

Approach:

- A comprehensive causality inference workbench
- Rephrasing some of the algorithmic calculation of causality as satisfiability queries which allows us to reuse the optimization power built in SAT and ILP solvers

Flavors of Causality

Spectrum-Based Fault Localization

Model-Based Diagnosis

Granger Causality

Halpern-Pearl Causality

- Definition
- **SAT-based computation**
- ILP-based computation

The Three Layer Causal Hierarchy

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in$Y$?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?

Figure 1: The Causal Hierarchy. Questions at level i can only be answered if information from level i or higher is available.

SAT-based Approach: Introduction

- Idea: **Instead of iterating** through all possible $\overrightarrow{W}$ create a **Boolean formula** F whose **satisfiability** indicates whether AC2 holds. Similarly for AC3.
  - Use equivalence operator to represent the equations
  - To account for the W set, the formula ORs the equivalence with the original variable.
  - Comparing the original, and sat-solved values determines the set w

- Benefit: Take advantage of **highly efficient** SAT-solvers [1, 3, 24, 26]

- Recall **AC2.** There is a set $\overrightarrow{W}$ of variables in V and a setting $\overrightarrow{x}'$ of the variables in $\vec{X}$ such that if $(M,\vec{u}) \vDash (\overrightarrow{W} = \vec{w})$, then $(M,\vec{u}) \vDash [\vec{X} \leftarrow \vec{x}, \overrightarrow{W} \leftarrow \vec{w}]\neg\varphi$

- Such a formula F has to

– incorporate $\neg\varphi$, context $\vec{u}$ the modified setting $\overrightarrow{x'}$ for potential cause $\vec{X}$

– take all possible **variations** of $\overrightarrow{W}$ into account and

– keep all other **semantics** of the underlying **causal model** M unchanged.

Amjad IBRAHIM, Simon REHWALD, Alexander PRETSCHNER: Efficiently Checking Actual Causality with SAT Solving.
To appear in Dependable Systems Engineering (Marktoberdorf Summer School 2019), IOS Press, 2019

SAT-based Approach: AC2 Algorithm

Algorithm 4 Check whether AC2 holds using SAT

Input: causal model M, context $\langle U_1, \ldots, U_n \rangle = \langle u_1, \ldots, u_n \rangle$, effect φ, tentative cause $\langle X_1, \ldots, X_\ell \rangle = \langle x_1, \ldots, x_\ell \rangle$, evaluation $\langle V_1, \ldots, V_m \rangle = \langle v_1, \ldots, v_m \rangle$

function FULFILLSAC2$(M, \vec{U} = \vec{u}, \varphi, \vec{X} = \vec{x}, \vec{V} = \vec{v})$

Observed values of endogeneous variables

Values of exogeneous variables

    if $(M, \vec{u}) \models [\vec{X} \leftarrow \neg\vec{x}]\neg\varphi$ **then return** $\varnothing$

    else

Values of exogeneous variables remain unchanged

$$F := \neg\varphi \wedge \bigwedge_{i=1\ldots n} f(U_i = u_i) \wedge \bigwedge_{i=1\ldots m, \nexists j \bullet X_j = V_i} (V_i \leftrightarrow F_{V_i} \vee f(V_i = v_i)) \wedge \bigwedge_{i=1\ldots\ell} f(X_i = \neg x_i)$$

Flipped tentative cause

End. variables as defined by model or as observed

$$\text{with } f(Y = y) = \begin{cases} Y, & y = 1 \\ \neg Y, & y = 0 \end{cases}$$

        if $\langle U_1 = u_1, \ldots, U_n = u_n, V_1 = v'_1, \ldots, V_m = v'_m \rangle = SAT(CNF(F))$ **then**

            $\vec{W} := \langle W_1, \ldots, W_s \rangle$ s.t. $\forall i \forall j \bullet (i \neq j \Rightarrow W_i \neq W_j) \wedge (W_i = V_j \Leftrightarrow v'_j = v_j)$

            return $\vec{W}$

        else return *null*

        end if

    end if

end function

Contains those end. variables whose value
is the same as observed, i.e., not flipped

AC3

Analysis of the satisfying assignments of *G*:
If we find a satisfying assignment for G, including the negation of the effect, such that at least one conjunct of the cause X =x takes on a value equal to

- its equation *or*
- its original value,

then this conjunct is **not a necessary part** of *X =x* so that AC2 is fulfilled.

Why? Because then X=x leads to both φ and $\neg\varphi$!

Checking AC3 (with ALL-SAT)

Algorithm 6 Check whether AC3 holds using SAT

Input: causal model M, context $\langle U_1, \dots, U_n \rangle = \langle u_1, \dots, u_n \rangle$, effect φ, tentative cause $\langle X_1, \dots, X_\ell \rangle = \langle x_1, \dots, x_\ell \rangle$, evaluation $\langle V_1, \dots, V_m \rangle = \langle v_1, \dots, v_m \rangle$

function FULFILLSAC3($M, \vec{U} = \vec{u}, \varphi, \vec{X} = \vec{x}, \vec{V} = \vec{v}$)

  if $\ell > 1 \wedge (M, \vec{u}) \models \varphi$ **then**

$$G := \neg\varphi \wedge \bigwedge_{i=1\dots n} f(U_i = u_i) \wedge \bigwedge_{i=1\dots m, \nexists j \bullet X_j = V_i} \left(V_i \leftrightarrow F_{V_i} \vee f(V_i = v_i)\right) \wedge \bigwedge_{i=1\dots\ell} X_i \vee \neg X_i$$

with $f(Y = y) = \begin{cases} Y, & y = 1 \\ \neg Y, & y = 0 \end{cases}$

    if $(A_G := \{\langle \vec{U} = \vec{u}, \vec{V} = \vec{v}' \rangle \mid \langle \vec{U} = \vec{u}, \vec{V} = \vec{v}' \rangle \in SAT(CNF(G))\}) \neq \varnothing$ **then**

      for all $\langle \vec{U} = \vec{u}, \vec{V} = \vec{v}' \rangle \in A_G$ **do**

        if $0 \le |\{j \in \{1, \dots, \ell\} | \exists i \bullet V_i = X_j \wedge v_i' \neq v_i \wedge v_i' \neq [\vec{V} \leftarrow \vec{v}'] F_{X_j}\}| < \ell$ **then**

  return 0

        end if

      end for

    end if

  end if

  return 1

end function

X_j must have been flipped

X_j=v_i' is an actual intervention, not a consequence of the model

All X_j must have been flipped for minimality

SAT-based Approach: AC3 without ALL_SAT

- Extend G to G'
  - With notions of non-minimality and non-emptiness

- UNSAT of G' entails that AC3 holds

Flavors of Causality

Spectrum-Based Fault Localization

Model-Based Diagnosis

Granger Causality

Halpern-Pearl Causality

Definition

SAT-based computation

ILP-based computation

The Three Layer Causal Hierarchy

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in Y?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing Intervening	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?

Figure 1: The Causal Hierarchy. Questions at level i can only be answered if information from level i or higher is available.

From SAT to ILP

- ILP can be used as a sat solver. Better: it can optimize the solution
- Researchers have done the transformation in the two directions
  - We will reuse our sat formulas
  - They already have the constraints we need
- Converting the formulas to ILP can happen at two levels:
  - Higher level: the level of F or G formulas
    - Formalize the equivalence as XNOR, then translate to linear constraints
  - CNF level [30]: Then we have clauses (disjunctions) that can be reduced to ILP constraints almost directly.
- Translation from SAT to ILP is standard:
  - Express $y=x1 \wedge x2$ as $0 \leq x_1+x_2-2*y \leq 1$
  - Express $y=x1 \vee x2$ as $0 \leq 2*y-x_1-x_2 \leq 1$
  - Express $y=\neg x$ as $y=1-x$
  - Express $y=x_1 \wedge ... \wedge x_n$ as $0 \leq x_1 + \ldots + x_n - n*y \leq n-1$
  - Express $y=x_1 \vee ... \vee x_n$ as $0 \leq n*y - x_1 - \ldots - x_n \leq n-1$

ILP Algorithm

1. Generate G formula
   a. Same as in SAT-based algorithm for AC3
   b. → CNF
2. Convert to ILP
   a. Using transformations from the literature
3. Create a distance measure
   a. The distance should be ≥1 and less or equal the size of X
4. Solve the program by minimizing the distance
   a. Testing with Gurobi [http://www.gurobi.com/]
5. Process results
   a. If model is feasible and optimal solution was found
      i. The distance indicates the size of the minimal cause
      ii. The values indicate which parts of the cause are required to be flipped
      iii. Inferring W is not discussed here

Benchmarked Models and Scenarios

12 different causal models: (5 causality literature, 1 attack tree, 2 fault trees, 4 artificial)

Artificial

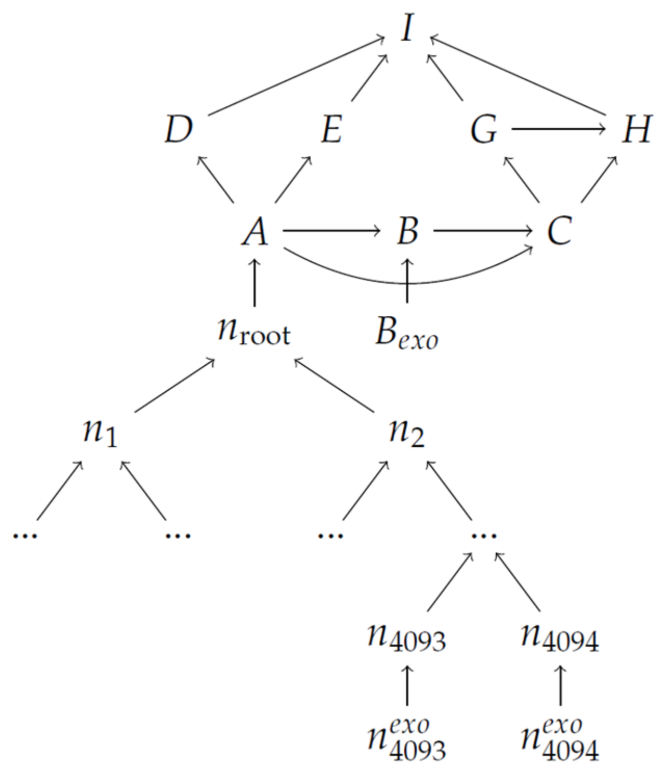


Figure 5.9.: Causal Graph of Abstract Causal Model 1 Combined with Binary Tree

Results

- Benchmarked on an Intel Core i7-4700HQ (2.40 GHz) with 4GB RAM (Windows 10)
- Framework: Java Microbenchmark Harness (JMH)
- SAT Solver: MiniSAT [3]

Benchmarked Models and Scenarios

Causal Model	Source	Number of Endogenous Variables
Rock-Throwing	[37, 40, 41]	5
Forest Fire (conjunctive & disjunctive)	[37, 40, 41]	3
Prisoners	[37, 41]	4
Assassin (first & second variant)	[37]	3
Railroad	[37]	4
Abstract Model 1 & 2	own example	8 & 3
Steal Master Key	industrial partner	36
Steal Master Key with Eight Attackers	industrial partner	91
Leakage in Subsea Production System	[15]	41
Leakage in Subsea Production System with Preemption	based on [15]	41
Binary Tree	own example	15 - 4095
Abstract Model 1 Combined with Binary Tree	own example	4103

Table 5.1.: Evaluated Causal Models

Representative Results

			Result					Execution Time (s)			Memory consumption (GB)		
Model	ID	$\lvert\vec{X}\rvert$	AC1	AC2	AC3	$\lvert\vec{W}\rvert$	$\lvert\vec{X}_{min}\rvert$	SAT	SAT_OPT	ILP	SAT	SAT_OPT	ILP
SMK	11	3	Y	Y	Y	0	3	0.013	0.0048	0.0024	0.008	0.0027	0.0026
BT_11	35	4	Y	N	Y	N/A	4	7.37	7.50	3.24	2.03	2.03	1.05
ABT	4	2	Y	Y	Y	4086	2	8.40	8.99	4.54	2.04	2.03	1.05
	5	3	Y	Y	N	4090	2	9.41	8.14	3.88	2.04	2.03	1.05
	6	10	Y	Y	N	4086	2	16.77	8.03	5.11	4.2	1.80	1.05
	7	11	N	Y	N	4086	2	26.29	8.07	5.26	4.18	1.80	1.05
	8	15	Y	Y	N	4086	2	N/A	8.10	5.108	N/A	1.81	1.05
	9	15	N	Y	N	4080	5	7412	7.87	5.05	9.5	1.80	1.05
	10	15	N	Y	N	4080	5	7301	8.23	5.27	9.5	1.80	1.05
	11	50	Y	Y	N	4079	5	N/A	8.55	4.80	N/A	2.04	1.05
	11	50	Y	Y	N	4068	11	N/A	8.21	4.76	N/A	2.04	1.05

Table 2: Discussed scenarios as part of the analysis

Bottom line

Some things will go wrong - need to cope with this. Hence accountability.

Monitoring and causal analysis.

Causal analysis at various levels: correlation, intervention, contrafactual.

Need for causal models. Reuse (or abuse) from various analysis tasks. Causal models necessarily incomplete.

HP logics for counterfactual reasoning. For binary models, efficient computations for answering queries possible in spite of NP.

References I

1 G. Audemard and L. Simon. “Predicting Learnt Clauses Quality in Modern SAT Solvers.” In: *IJCAI 2009, Proceedings of the 21st International Joint Conference on Artificial Intelligence, Pasadena, California, USA, July 11-17, 2009*. 2009, pp. 399–404.

2 I. R. Edwards. “Considerations on causality in pharmacovigilance.” In: *International Journal of Risk and Safety in Medicine* 24.1 (2012). cited By 6, pp. 41–54. DOI: 10.3233/JRS-2012-0552.

3 N. Eén and N. Sörensson. “An ESTensible SAT-solver.” In: *Theory and Applications of Satisfiability Testing, 6th International Conference, SAT 2003. Santa Margherita Ligure, Italy, May 5-8, 2003 Selected Revised Papers*. 2003, pp. 502–518. DOI: 10.1007/978-3-540-24605-3_37.

4 B. Fazzinga, S. Flesca, F. Furfaro, and L. Pontieri. “Online and offline classification of traces of event logs on the basis of security risks.” In: *J. Intell. Inf. Syst.* 50.1 (2018), pp. 195–230. DOI: 10.1007/s10844-017-0450-y.

5 J. Feigenbaum, J. A. Hendler, A. D. Jaggard, D. J. Weitzner, and R. N. Wright. “Accountability and deterrence in online life.” In: *Web Science 2011, WebSci ’11, Koblenz, Germany - June 15 - 17, 2011*. 2011, 7:1–7:7. DOI: 10.1145/2527031.2527043.

References II

[6] J. Feigenbaum, A. D. Jaggard, and R. N. Wright. "Towards a formal model of accountability." In: *2011 New Security Paradigms Workshop, NSPW '11, Marin County, CA, USA, September 12-15, 2011*. 2011, pp. 45–56. DOI: 10.1145/2073276.2073282.

7 I. Freckelton, ed. *Causation in law and medicine*. Aldershot: Ashgate/Dartmouth, 2002. ISBN: 0-7546-2204-5.

8 A. Groce. "Error Explanation with Distance Metrics." In: *Tools and Algorithms for the Construction and Analysis of Systems, 10th International Conference, TACAS 2004, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2004, Barcelona, Spain, March 29 - April 2, 2004, Proceedings*. 2004, pp. 108–122. DOI: 10.1007/978-3-540-24730-2_8.

9 A. Groce, S. Chaki, D. Kroening, and O. Strichman. "Error explanation with distance metrics." In: *STTT* 8.3 (2006), pp. 229–247. DOI: 10.1007/s10009-005-0202-0.

[10] J. Y. Halpern. "A Modification of the Halpern-Pearl Definition of Causality." In: *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*. 2015, pp. 3022–3033.

[11] J. Y. Halpern. *Actual causality*. Cambridge, Massachussetts: The MIT Press, 2016. ISBN: 978-0-262-03502-6; 0-262-03502-2.

References III

[12] J. Y. Halpern and J. Pearl. “Causes and Explanations: A Structural-Model Approach - Part I: Causes.” In: *UAI ’01: Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence, University of Washington, Seattle, Washington, USA, August 2-5, 2001*. 2001, pp. 194–202.

[13] J. Y. Halpern and J. Pearl. “Causes and Explanations: A Structural-Model Approach. Part I: Causes.” In: *The British Journal for the Philosophy of Science* 56.4 (2005), pp. 843–887. DOI: 10.1093/bjps/axi147. eprint: /oup/backfile/content_public/journal/bjps/56/4/10.1093/bjps/axi147/2/axi147.pdf.

[14] J. Y. Halpern and J. Pearl. “Causes and Explanations: A Structural-Model Approach - Part II: Explanations.” In: *Proceedings of the Seventeenth International Joint Conference on Artificial Intelligence, IJCAI 2001, Seattle, Washington, USA, August 4-10, 2001*. 2001, pp. 27–34.

[15] J. Y. Halpern and J. Pearl. “Causes and Explanations: A Structural-Model Approach. Part II: Explanations.” In: *The British Journal for the Philosophy of Science* 56.4 (2005), pp. 889–911. DOI: 10.1093/bjps/axi148. eprint: /oup/backfile/content_public/journal/bjps/56/4/10.1093/bjps/axi148/2/axi148.pdf.

[16] D. Hume. *A Treatise of Human Nature*. Oxford University Press, 1738.

[17] M. Jose and R. Majumdar. “Cause clue clauses: error localization using maximum satisfiability.” In: *Proceedings of the 32nd ACM SIGPLAN Conference on Programming Language Design and Implementation, PLDI 2011, San Jose, CA, USA, June 4-8, 2011*. 2011, pp. 437–446. DOI: 10.1145/1993498.1993550.

References IV

[18] S. Kacianka, F. Kelbert, and A. Pretschner. "Towards a Unified Model of Accountability Infrastructures." In: *Proceedings First Workshop on Causal Reasoning for Embedded and safety-critical Systems Technologies, CREST@ETAPS 2016, Eindhoven, The Netherlands, 8th April 2016*. 2016, pp. 40–54. DOI: 10.4204/EPTCS.224.5.

[19] S. Khan and S. Parkinson. "Causal Connections Mining Within Security Event Logs." In: *Proceedings of the Knowledge Capture Conference, K-CAP 2017, Austin, TX, USA, December 4-6, 2017*. 2017, 38:1–38:4. DOI: 10.1145/3148011.3154476.

[20] S. Kleinberg and G. Hripcsak. "A review of causal inference for biomedical informatics." In: *Journal of Biomedical Informatics* 44.6 (2011), pp. 1102–1112. DOI: 10.1016/j.jbi.2011.07.001.

[21] R. Künnemann, I. Esiyok, and M. Backes. "Automated Verification of Accountability in Security Protocols." In: *CoRR* abs/1805.10891 (2018). arXiv: 1805.10891.

[22] D. Lewis. "Causation." In: *Journal of Philosophy* 70.17 (1973), pp. 556–567. DOI: 10.2307/2025310.

[23] M. S. Moore. *Causation and responsibility : an essay in law, morals, and metaphysics*. Oxford: Oxford Univ. Press, 2009. ISBN: 978-0-19-925686-0.

[24] M. W. Moskewicz, C. F. Madigan, Y. Zhao, L. Zhang, and S. Malik. "Chaff: Engineering an Efficient SAT Solver." In: *Proceedings of the 38th Annual Design Automation Conference*. DAC '01. Las Vegas, Nevada, USA: ACM, 2001, pp. 530–535. ISBN: 1-58113-297-2. DOI: 10.1145/378239.379017.

References V

[25] C. H. Papadimitriou and M. Yannakakis. “The Complexity of Facets (and Some Facets of Complexity).” In: *J. Comput. Syst. Sci.* 28.2 (1984), pp. 244–259. DOI: 10.1016/0022-0000(84)90068-0.

[26] J. P. M. Silva and K. A. Sakallah. “GRASP - a new search algorithm for satisfiability.” In: *ICCAD*. 1996, pp. 220–227. DOI: 10.1109/ICCAD.1996.569607.

[27] L. Traeger. *Der Kausalbegriff im Straf- und Zivilrecht : zugleich ein Beitrag zur Auslegung des BGB*. Marburg: Elwert, 1904.

[28] J. Williamson. “Handbook of Philosophical Logic.” In: ed. by D. Gabbay and F. Guenthner. Dordrecht: Springer Netherlands, 2007. Chap. Causality, pp. 95–126. ISBN: 978-1-4020-6324-4. DOI: 10.1007/978-1-4020-6324-4_2.

[29] L. B. L. Wittemans, L. A. Lotta, and C. Langenberg. “Prioritising Risk Factors for Type 2 Diabetes: Causal Inference through Genetic Approaches.” In: *Current Diabetes Reports* 18.7 (May 2018), p. 40. ISSN: 1539-0829. DOI: 10.1007/s11892-018-1009-1.

[30] Li, Ruiming, Dian Zhou, and Donglei Du. "Satisfiability and integer programming as complementary tools." Proceedings of the 2004 Asia and South Pacific design automation conference. IEEE Press, 2004.