A PEARL PEARL: how to teach an old AI new tricks

NADA AMIN [1], WILLIAM E. BYRD [2], JOSEPH A. COTTAM [3], MARC-ANTOINE PARENT [4], JEREMY D. ZUCKER [3]

1. HARVARD UNIVERSITY, USA; 2. UNIVERSITY OF ALABAMA, USA; 3. PACIFIC NORTHWEST NATIONAL LABORATORY, USA; 4. SOLUTIONS CONVERSENCE INC

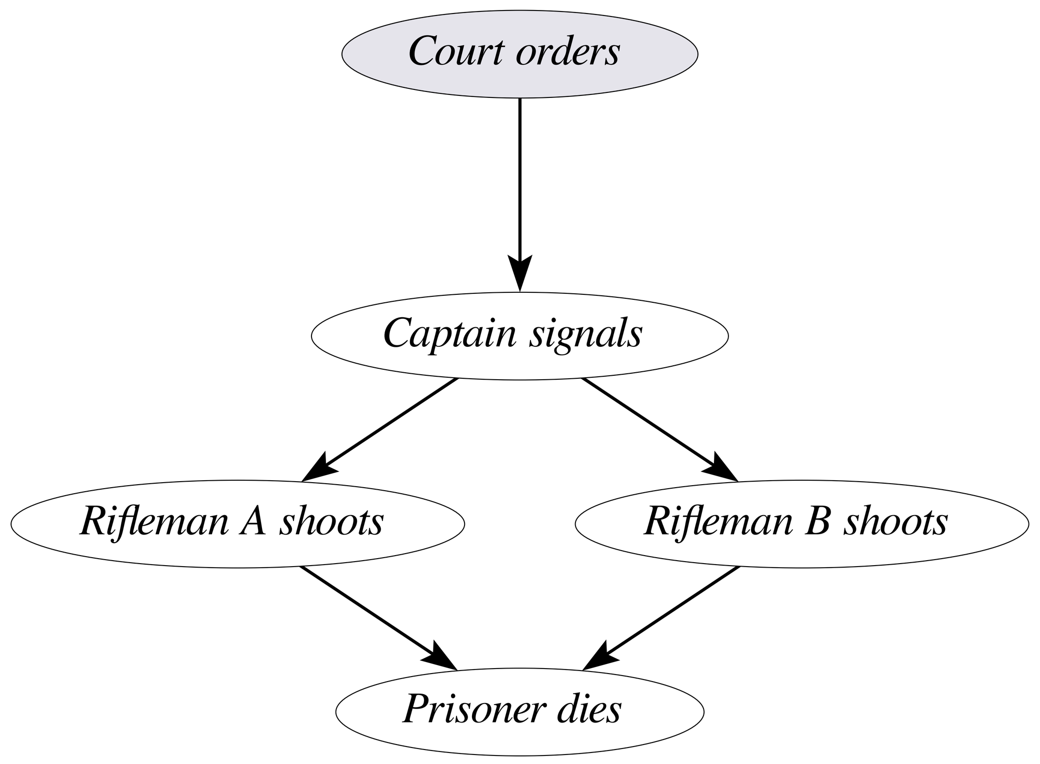
William E. Byrd's work was supported by the National Center For Advancing Translational Sciences of the National Institutes of Health under Award Number OT2TR003435. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Joseph A. Cottam and Jeremy D. Zucker's work was supported by the Data-Model Convergence Initiative, a component of the Laboratory Directed Research and Development Program at Pacific Northwest National Laboratory, a multiprogram national laboratory operated by Battelle for the U.S. Department of Energy under Contract DE-AC05-76RL01830.

ABSTRACT

Causal reasoning is a basic component of human intelligence. However, causal information has proven difficult for AI and machine learning applications to incorporate. This omission hinders such systems as they cannot take advantage of this important information during reasoning or explanation. Recent advances in formal causal reasoning open new possibilities to include causal reasoning in such systems. We show that causal reasoning can be represented in a Truth Maintenance Systems (TMS), and thus enable expert systems to reason causally. The integration is complete (demonstrated by the ability to answer all six 'Firing Squad' questions), simple (only 200 lines of Common Lisp), and practical (through an application to medical diagnosis). This paper outlines the implementation and shows how to use a TMS augmented with causal reasoning.

LEVEL 1: MONOTONIC LOGIC IS ALL YOU NEED



Prediction If rifleman A did not shoot, then prisoner is alive:

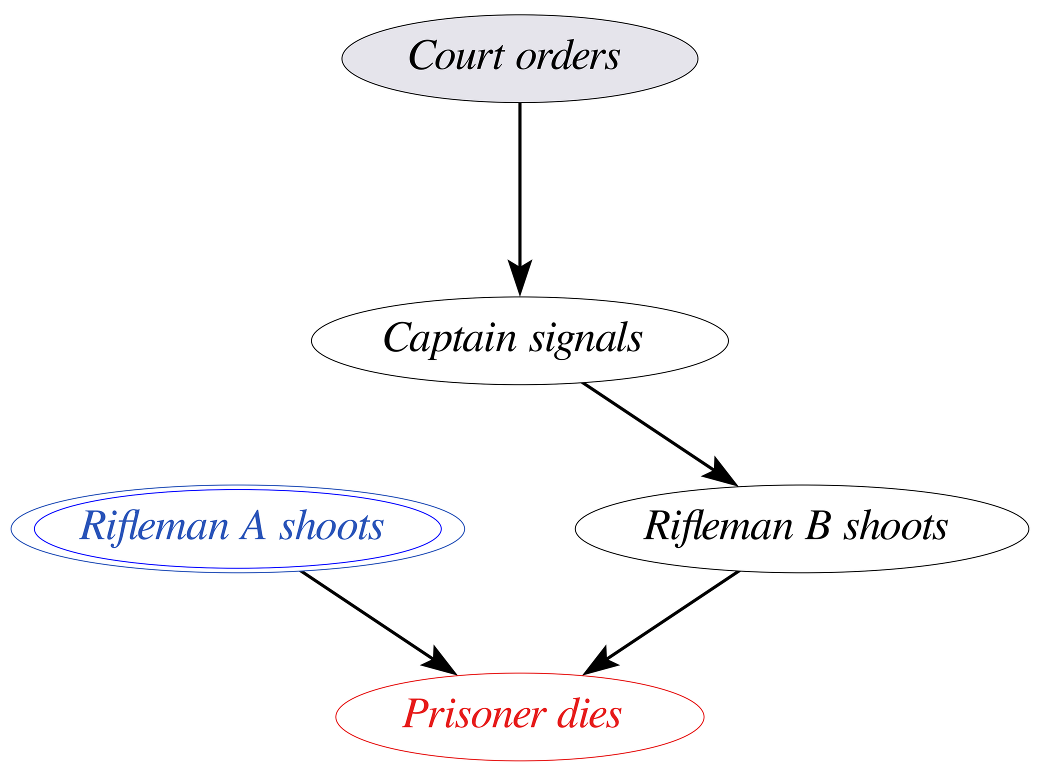
$$A = \neg a \implies D = \neg d$$

```
(enable-assumption A :FALSE)
(why-node D)
```

Prisoner dies is FALSE because

Rifleman A shoots is FALSE
Rifleman B shoots is FALSE
Rifleman A or B shoots
=> Prisoner dies is TRUE

LEVEL 2: ACTION THROUGH RETRACTION



If the Captain gave no signal, but Rifleman A decides to shoot, then the prisoner will be dead, and Rifleman B will not shoot:

$$C_{\neg A}(U = \neg u) = \neg c \implies D_{A=\neg a} = d \wedge B_{A=\neg a} = \neg b$$

```
(retract-assumption C=>A)
(enable-assumption C=>A :FALSE)
(enable-assumption A :TRUE)
(enable-assumption C :FALSE)
(why-node D)
```

Prisoner dies is TRUE because

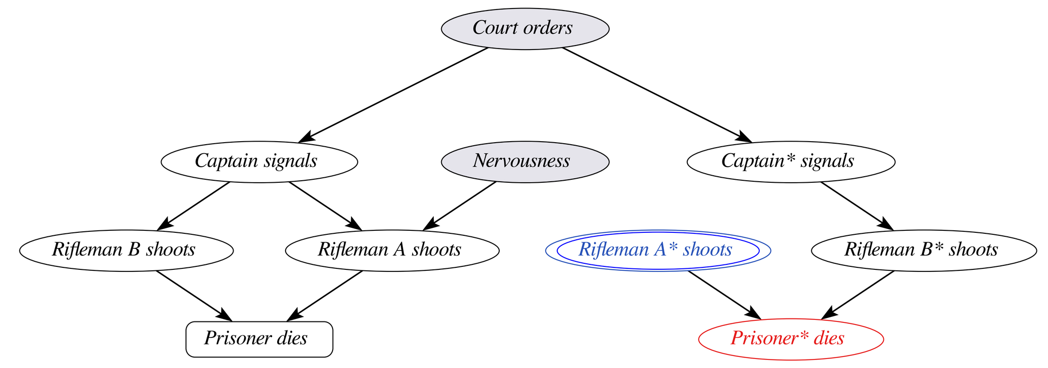
Rifleman A shoots is TRUE
Rifleman A or B shoots
=> Prisoner dies is TRUE

```
(why-node B)
```

Rifleman B shoots is FALSE because

Court orders is FALSE
Court orders => Captain signals is TRUE
Captain signals => Rifleman B shoots is TRUE

LEVEL 3: CAUSAL CRANK FOR PROBABILISTIC COUNTERFACTUALS



Suppose now that the Court gives an order with probability $P(U = u) = p$, and we also introduce the possibility that rifleman A has a twitchy finger W, and sometimes shoots regardless of the Captain's signal, with probability $P(W = w) = q$. We wish to know the probability that the Prisoner would be alive if Rifleman A had not shot, given that the Prisoner is dead:

$$P(D_{do(A=\neg a)} = \neg d | D = d)$$

```
(setq *causal* (make-causal :title "riflemen"
     :graph '(
          (Court-orders .  Captain-signals)
          (Captain-signals .  RiflemanA-shoots)
          (Captain-signals .  RiflemanB-shoots)
          (RiflemanA-shoots .  Prisoner-dies)
          (RiflemanB-shoots .  Prisoner-dies)
          (A-is-nervous .  RiflemanA-shoots)
     )
     :priors '(
          (Court-orders .  0.6)
          (A-is-nervous .  0.7)
     )
     :symbolic-priors '(
          (Court-orders .  p)
          (A-is-nervous .  q)
     )
     :given 'Prisoner-dies
     :intervention '(:NOT RiflemanA-shoots)
     :outcome '(:NOT Prisoner-dies)
))
(causal-crank *causal*)
```

Outcome probability: 0.32.

COMPUTING COUNTERFACTUALS BY COMPOSING TMS'S

Abduction We first construct a factual Assumption-based Truth Maintenance System (ATMS) from the causal graph. We then use the `:prior` variable-value assignments to create the possibility space as a joint distribution over all combinations of values of exogenous variables. To obtain the posterior distribution over the exogenous variables, we then filter and renormalize the original possibility space by considering only the possibilities where the `:given` event holds.

Action We then construct the counterfactual graph by severing the incoming links to the `:intervention` nodes. Next, we construct a distinct counterfactual ATMS for the counterfactual graph.

Prediction Finally, we port the posterior distribution over exogenous variables to the counterfactual world. In the counterfactual world, we then compute the `:outcome` probability. To find the probability of an event, we sum the probability of worlds where that event occurs, as computed by ATMS inference. This calculation works for any space with disjoint outcomes.

CONCLUSION AND REMAINING CHALLENGES

In real systems, the fully-specified structural causal model is rarely known, so much of causal inference is about inferring properties of the underlying SCM given partial observational and interventional information. Nonetheless, in biological systems, domain knowledge often allows for the specification of functional forms that can be parameterized from partial information.