Causal Knowledge Graph Explainability using Interventional and Counterfactual reasoning

KGC- Workshop for Imperfect reasoning

Utkarshani Jaimini
Ph.D. Candidate
Artificial Intelligence Institute
University of South Carolina

AI INSTITUTE #AIISC
UNIVERSITY OF SOUTH CAROLINA

Motivation

Current AI approaches rely on statistical correlations that are often spurious and can’t be explained.

The CausalKG platform delivers unparalleled explainability approach that works for AI-powered decision making.

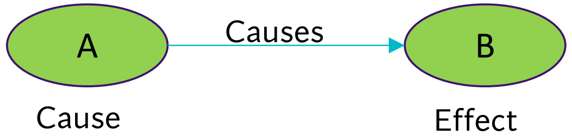
Causality

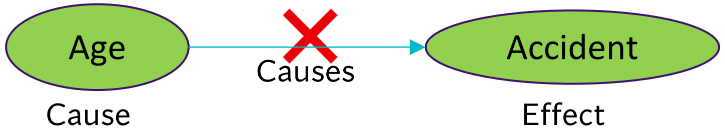
Causality is a relationship “A” causes “B”

Causality is at the core of everything we *see, do,* and *imagine*.

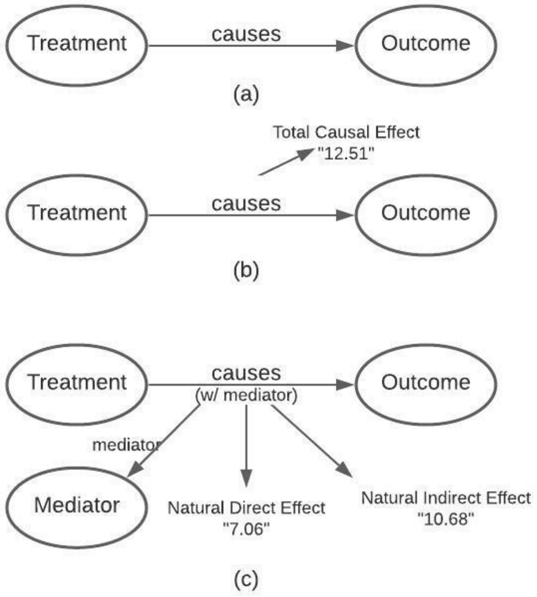
Correlation is not Causation

- Younger drivers have high probability of being in an accident
  - Does not imply younger drivers cause accidents





Causality Representation



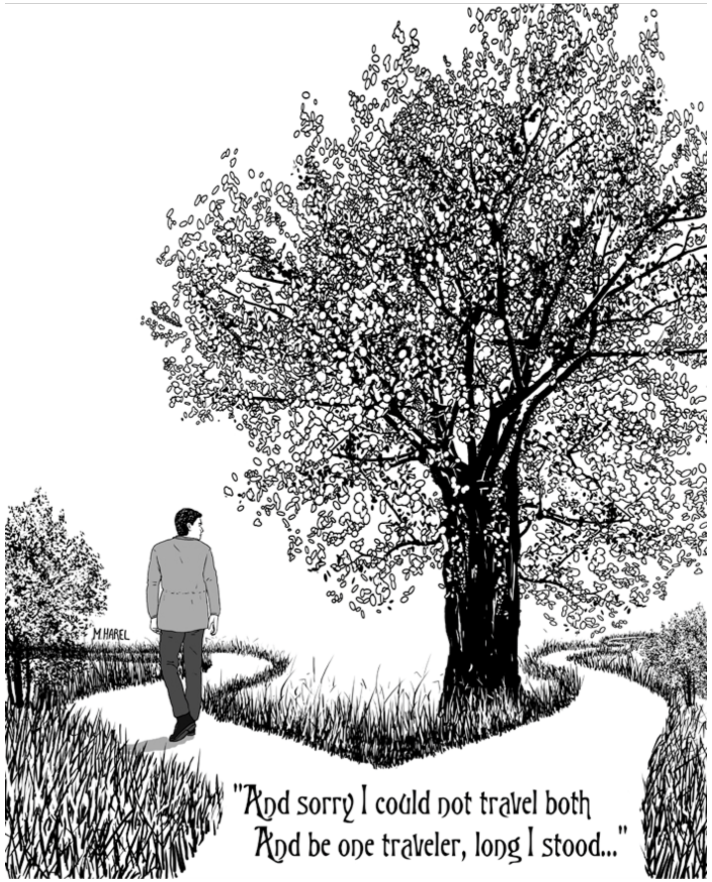
(a) Causal Representation as a single cause-effect relation.
(b), (c) Causality as a complex representation of causal effect associated with the different pathways.

Counterfactuals

WHAT If? scenarios

- Human mind has an ability to conceive alternative, nonexistent worlds

- We can see what might have happened
  - Imagine, be prepared and act in counterfactuals scenarios


M.HAREL
"And sorry I could not travel both
And be one traveler, long I stood..."

Counterfactuals

(The path not taken by Robert Frost)

http://danamackenzie.com/amazing-science-art/

Counterfactuals in AI

- It is not the instruction that "Do not vacuum"
- Vacuum cleaners make noise, noise wakes people up, and that makes people unhappy
- Understand *cause-and-effect* from *counterfactual relations* encoded in "You shouldn't have."

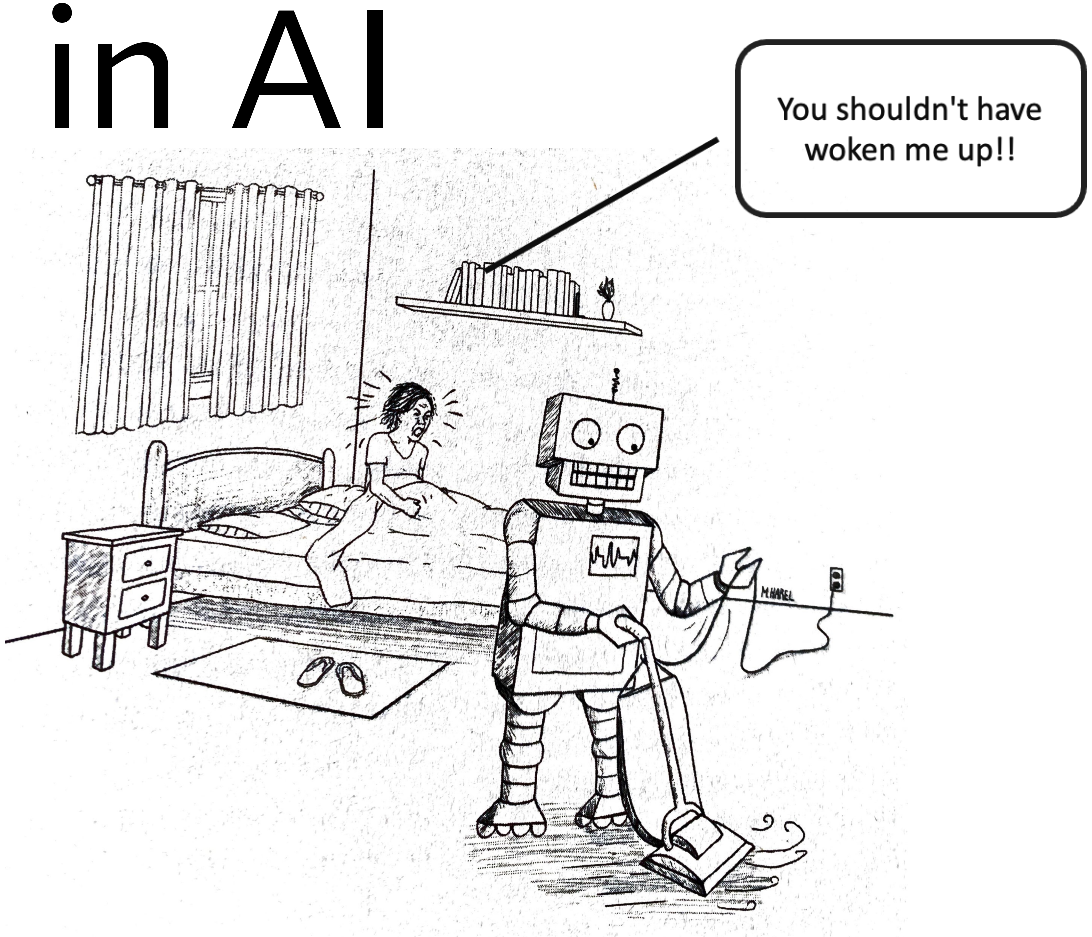


Figure: A smart robot contemplating the causal ramification of his/her action. (Source: Book of Why, Drawing by Maayan Harel)

http://danamackenzie.com/amazing-science-art/

Ladder of Causation

Dr. Judea Pearl

Future Hybrid AI

Current AI Approach

Association based on the observational data

IMAGINING
DOING
SEEING
M. HAREL

3. COUNTERFACTUALS, $P(y_x \mid x', y')$

ACTIVITY: Imagining, Retrospection, Understanding

QUESTIONS: What if I had done ....? What if I had acted differently? Was it X that caused Y? What if X had not occurred?

EXAMPLES: Was it the Pedestrian crossing sign which led to the vehicle's stop?

2. INTERVENTION, $P(y \mid do(x), z)$

ACTIVITY: Doing, Intervening

QUESTIONS: What if I do ....? What if I do X? What would Y be if I do X?

EXAMPLES: What if the pedestrian crossing sign is off?

1. ASSOCIATION, $P(y \mid x)$

ACTIVITY: Seeing, Observing

QUESTIONS: What if I see....? How would seeing X change my belief in Y?

EXAMPLES: What is the probability of a vehicle stopping if there is a pedestrian crossing the street?

Figure: Judea Pearl's Ladder of Causation

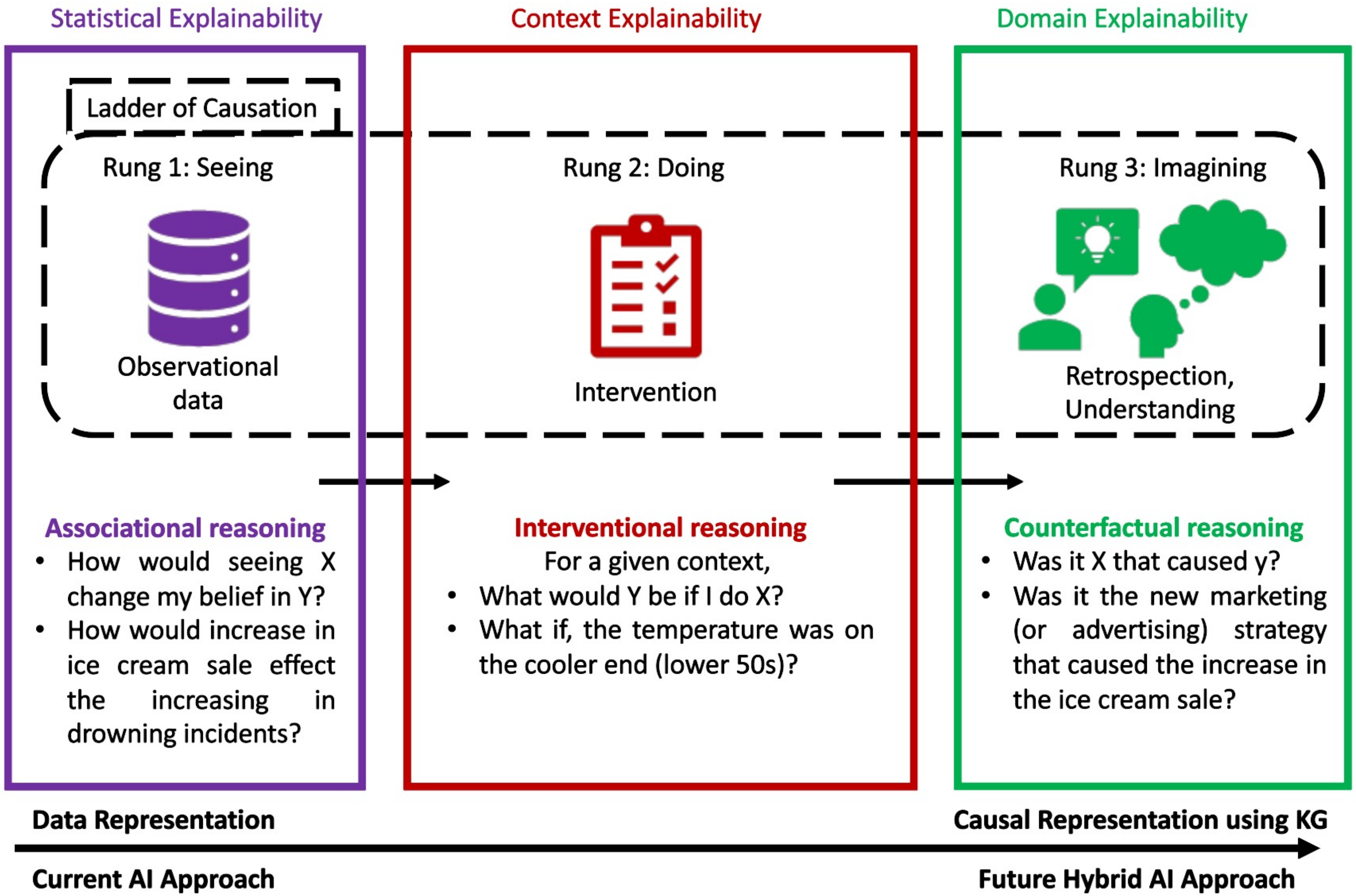
http://danamackenzie.com/amazing-science-art/

Causality for Explainability

Representation of causality in AI systems using knowledge-graph based approach is needed for better explainability, and support for intervention and counterfactuals, leading to improved understanding of AI systems by humans.

Explainability

- **Statistical explainability** generates an explanation for the co-occurrence of a given phenomenon based on the statistical (or associational) methods such as correlations.
- **Context explainability** is a means to generate a human-understandable explanation taking the context information of a given observation into account.
- **Domain explainability** explains the underlying causal relations using observational data, domain knowledge, and counterfactual reasoning.



Statistical Explainability

Ladder of Causation

Rung 1: Seeing

Observational data

Associational reasoning

- How would seeing X change my belief in Y?
- How would increase in ice cream sale effect the increasing in drowning incidents?

Context Explainability

Rung 2: Doing

Intervention

Interventional reasoning

For a given context,

- What would Y be if I do X?
- What if, the temperature was on the cooler end (lower 50s)?

Domain Explainability

Rung 3: Imagining

Retrospection, Understanding

Counterfactual reasoning

- Was it X that caused y?
- Was it the new marketing (or advertising) strategy that caused the increase in the ice cream sale?

Data Representation

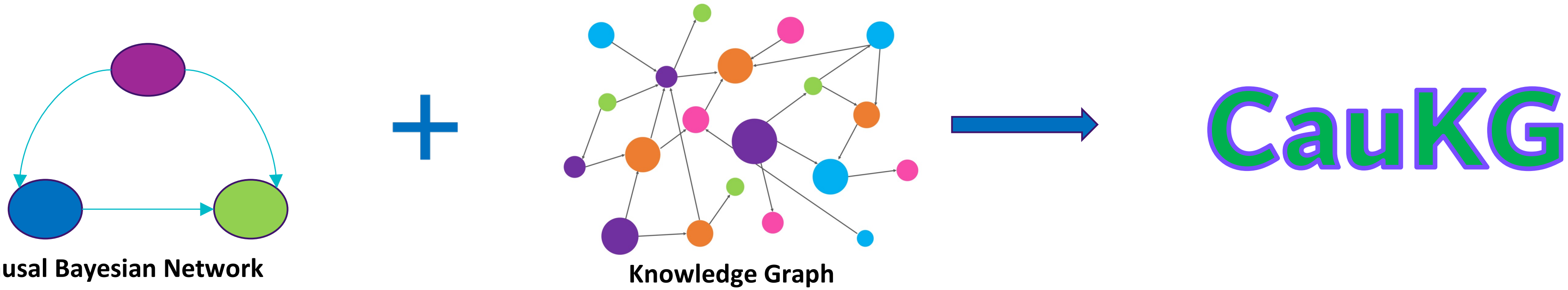
Causal Representation using KG

Current AI Approach

Future Hybrid AI Approach

Figure inspired from the Ladder of causation (Dr. Judea Pearl),
and Ladder of thinking (Notger Heinz)

Causal Knowledge Graph



Pros

Models the causal relations between the corresponding nodes

Cons

Different experts can suggest different causal model

Pros

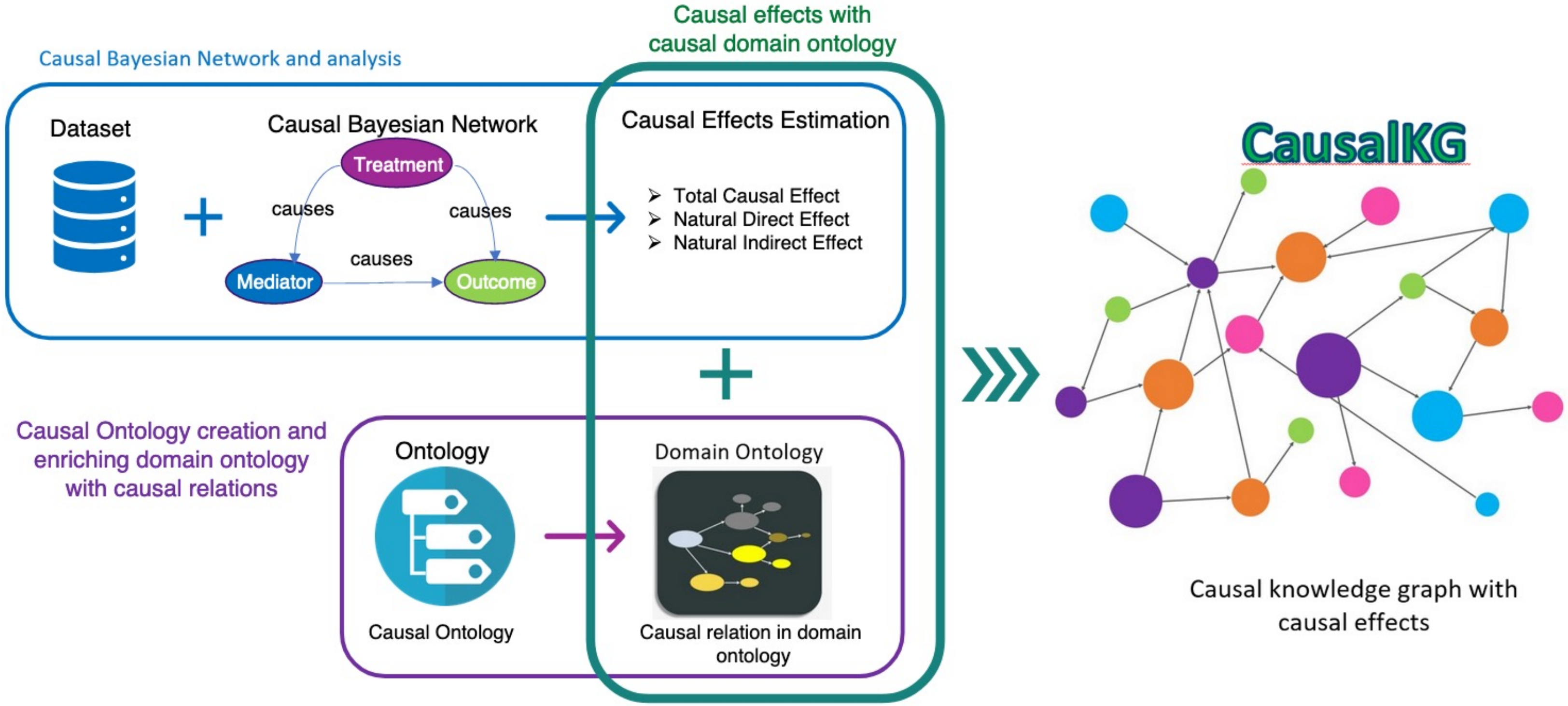
Extensive domain knowledge representation

Cons

Lacking causal knowledge representation techniques

Representation of causal facts from the dataset using causal Bayesian networks, causal ontology and knowledge representation technique for downstream tasks

Causal Knowledge Graph

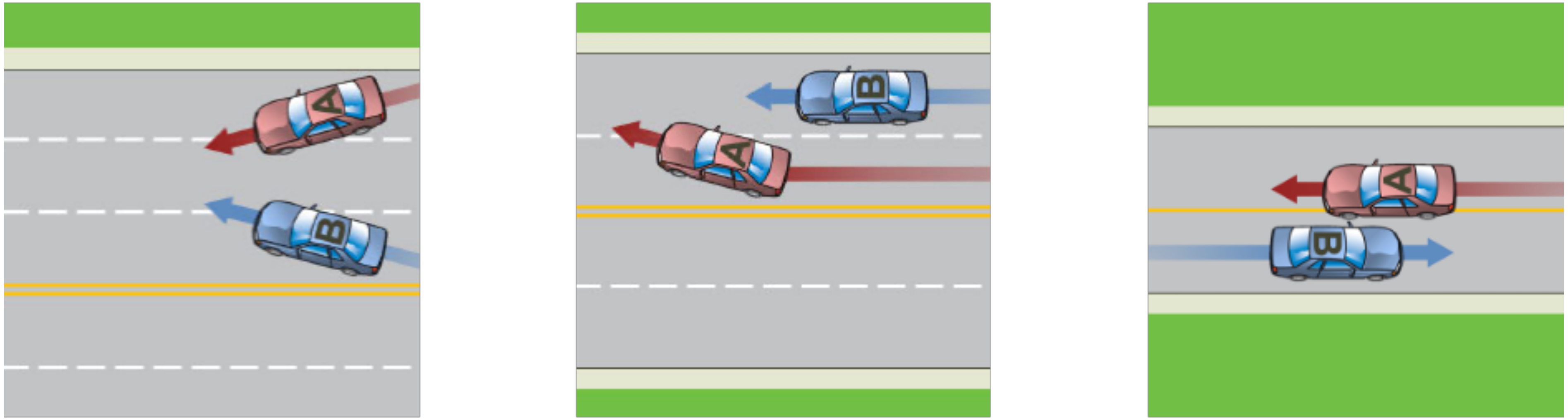

Causal Bayesian Network and analysis
Dataset
Causal Bayesian Network
Treatment
causes
causes
causes
Mediator
Outcome
Causal effects with causal domain ontology
Causal Effects Estimation
➢ Total Causal Effect
➢ Natural Direct Effect
➢ Natural Indirect Effect
Causal Ontology creation and enriching domain ontology with causal relations
Ontology
Causal Ontology
Domain Ontology
Causal relation in domain ontology
CausalKG
Causal knowledge graph with causal effects

CausalKG framework consists of three main steps, 1) a **CBN** and a domain-specific observational dataset, 2) **Causal Ontology** creation and enriching the domain ontology with causal relationships, and 3) Estimating the **causal effects** of the treatment, mediator, and outcome variable in the domain for a given context.

Use Case

Driving Scene

Driving: Lane change



(a)

(b)

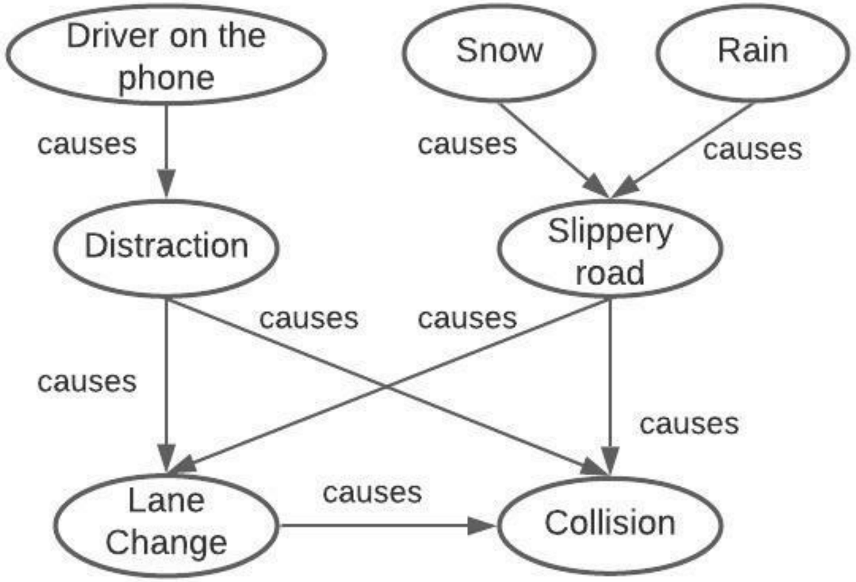
(c)

Three possible risky lane change scenes. The above situations might arise either due to a distracted driver, blind spot, or slippery road due to snow or rain (weather condition).

https://www.icbc.com/claims/crash-responsibility-fault/crash-examples/Pages/lane-change-crash.aspx

Causal Bayesian Network

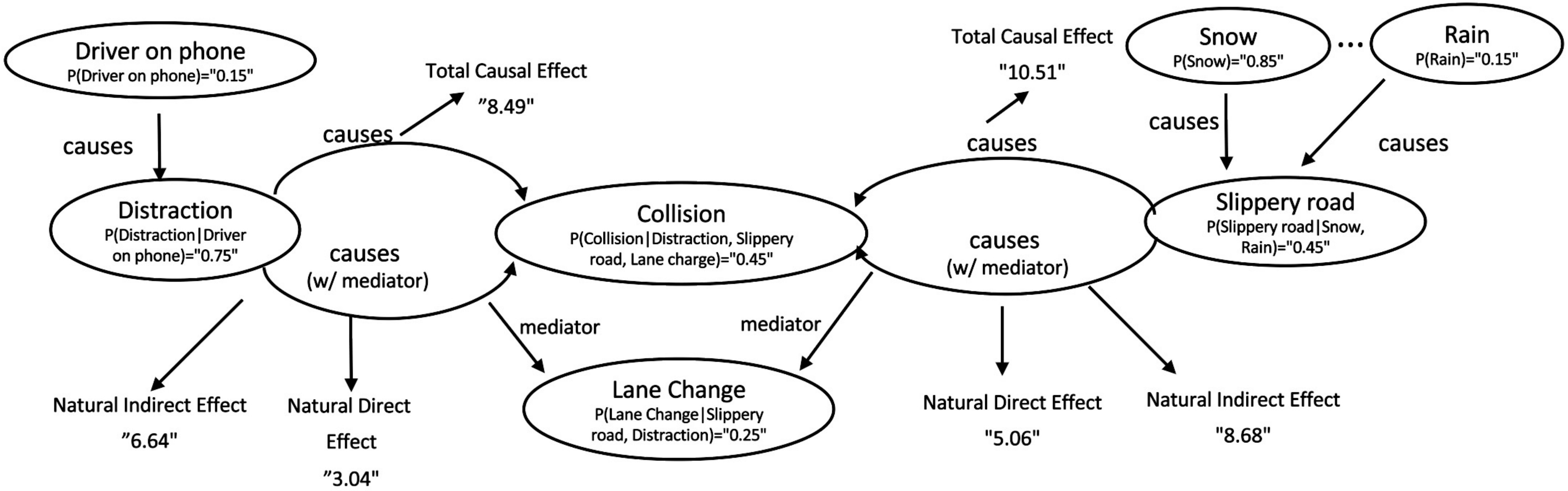
For driving lane change scenario



Competency Questions

- Total causal effect (Basic cause): How would the driver's distraction (or slippery road) effect the occurrence of a vehicle collision?
- Natural direct effect (unplanned or direct cause): What if the vehicle fails to identify the passing vehicle in the adjacent lane (what if there is a blind spot), how would it effect the vehicle's collision? In this scenario, the lane change is not due to the distraction (or slippery road), but the collision is due to the distraction. ›
- Natural indirect effect (unsafe act or indirect cause): What if, there is a lane change due to the distraction (or slippery road), how would it effect the vehicle's collision? In this scenario, the lane change is due to distraction (or losing control over the vehicle under risky situations). However, the collision is not due to distraction.

Snapshot of CausalKG



Snapshot of CausalKG for collisions due to driver distraction or slippery road scenarios. Each node in the CausalKG is a concept in KG and is associated with a conditional probability estimated using the CBN. The edges between the nodes represent the causal relationships between the concepts.

Future

- With CauKG we can utilize *knowledge-infused learning* to inject causal knowledge into machine learning models
- Integrating the CausalKG with the autonomous driving scene, and healthcare knowledge graph
- Learning causal knowledge hyper relational graph embeddings

Thank You

Acknowledgement:

Dr. Amit Sheth (Advisor)

Dr. Cory Henson

National Science Foundation Award #2133842 “EAGER: Advancing Neuro-symbolic AI with Deep Knowledge infused Learning,” and Award #2119654, “RII Track 2 FEC: Enabling Factory to Factory Networking for Future Manufacturing.”