

# Workshop Week 8: Playing resistance

CITS3001 Algorithms, Agents and Artificial  
Intelligence

---

Tim French  
Dept of Computer Science and Software Engineering  
The University of Western Australia

---

2021, Semester 2

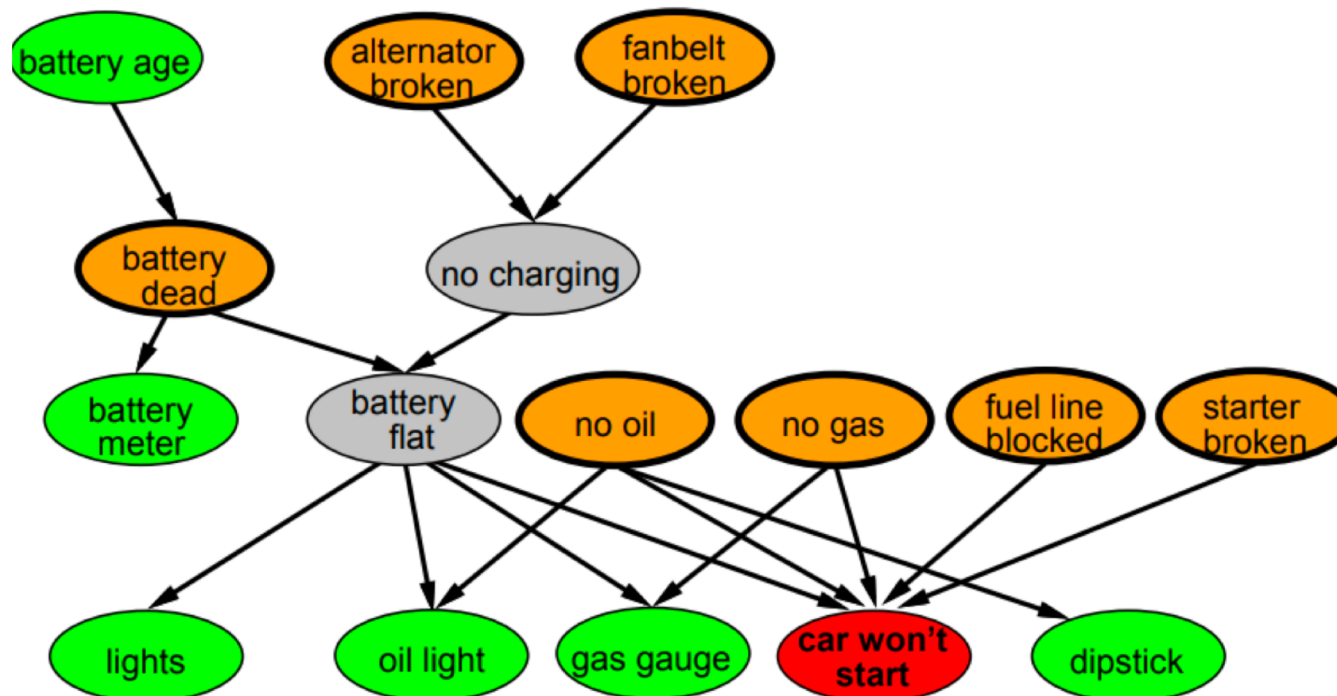
# Modeling Uncertainty

1) We will first review the lecture material on learning under uncertainty.

Initial evidence: car won't start

Testable variables (green), "broken, so fix it" variables (orange)

Hidden variables (gray) ensure sparse structure, reduce parameters



# Agent Heuristics for The Resistance.



- What is the size of the state space for the game?
- How can we represent an agent's policy in The Resistance?

# Heuristics to MDPs

- Assuming heuristics for the opposing players can we turn The Resistance strategy selection into a Markov Decision Process?
- How can we use sampling to help us choose the best action.

# Uncertainty

- 1) There are only limited types of uncertainty in The Resistance :
- 2) The Resistance does not know who the spies are.
- 3) You don't know the policy other players are using.

However there are ways to make inferences from players actions:

- Assuming players are using a policy, you can infer a posterior distribution of who the spies are.
- Confidence comes from evidence: the more you see other agents vote, the more information you have.
- Policies will change between rounds: you expect spies to be conservative early, and aggressive later.

Bayes Theorem is useful for maintaining a probability distribution.

In a short game, it is probably best to assume a static policy for other players, although you could make broad assumptions. However, if you have access to a corpus of games, you can make assumptions about what policies are most likely to be used.

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

Diagram illustrating Bayes' Theorem with handwritten annotations:

- $P(A|B)$ : THE PROBABILITY OF "A" BEING TRUE GIVEN THAT "B" IS TRUE
- $P(B|A)$ : THE PROBABILITY OF "B" BEING TRUE GIVEN THAT "A" IS TRUE
- $P(A)$ : THE PROBABILITY OF "A" BEING TRUE
- $P(B)$ : THE PROBABILITY OF "B" BEING TRUE