Under the Hood
Probabilistic Laws of Robotics and Networks of AIs

(At least three) Technical areas of interest

- Reasoning under Uncertainty
- Lookahead (projecting into the future)
- Communication among AIs, directly or through a central authority

Reasoning under Uncertainty

Current situation

Swerve right

2 dead, 3 minor injuries (Pr = 0.1) (C = 100)

1 dead, 4 minor injuries (Pr = 0.4) (C = 80)

0 dead, 1 major injury,
4 minor injuries Pr = (0.5) (C = 10)

Current situation

Continue

3 dead, 2 minor injuries (Pr = 0.8) (C = 150)

1 dead, 4 minor injuries (Pr = 0.2) (C = 80)

Current situation

Swerve left

3 dead, 1 major, 1 minor injuries (Pr = 0.3) (C = 160)

2 dead, 3 minor injuries (Pr = 0.5) (C = 100)

1 dead, 0 injuries Pr = (0.2) (C = 70)

Swerve Right (expected cost): (0.1 * 100) + (0.4 * 80) + (0.5 * 10) = 47 Choose "Swerve Right"

Continue (expected cost): (0.8 * 150) + (0.2 * 80) = 136

Swerve Left (expected cost): (0.3 * 160) + (0.5 * 100) + (0.2 * 70) = 112

Issues with uncertainty

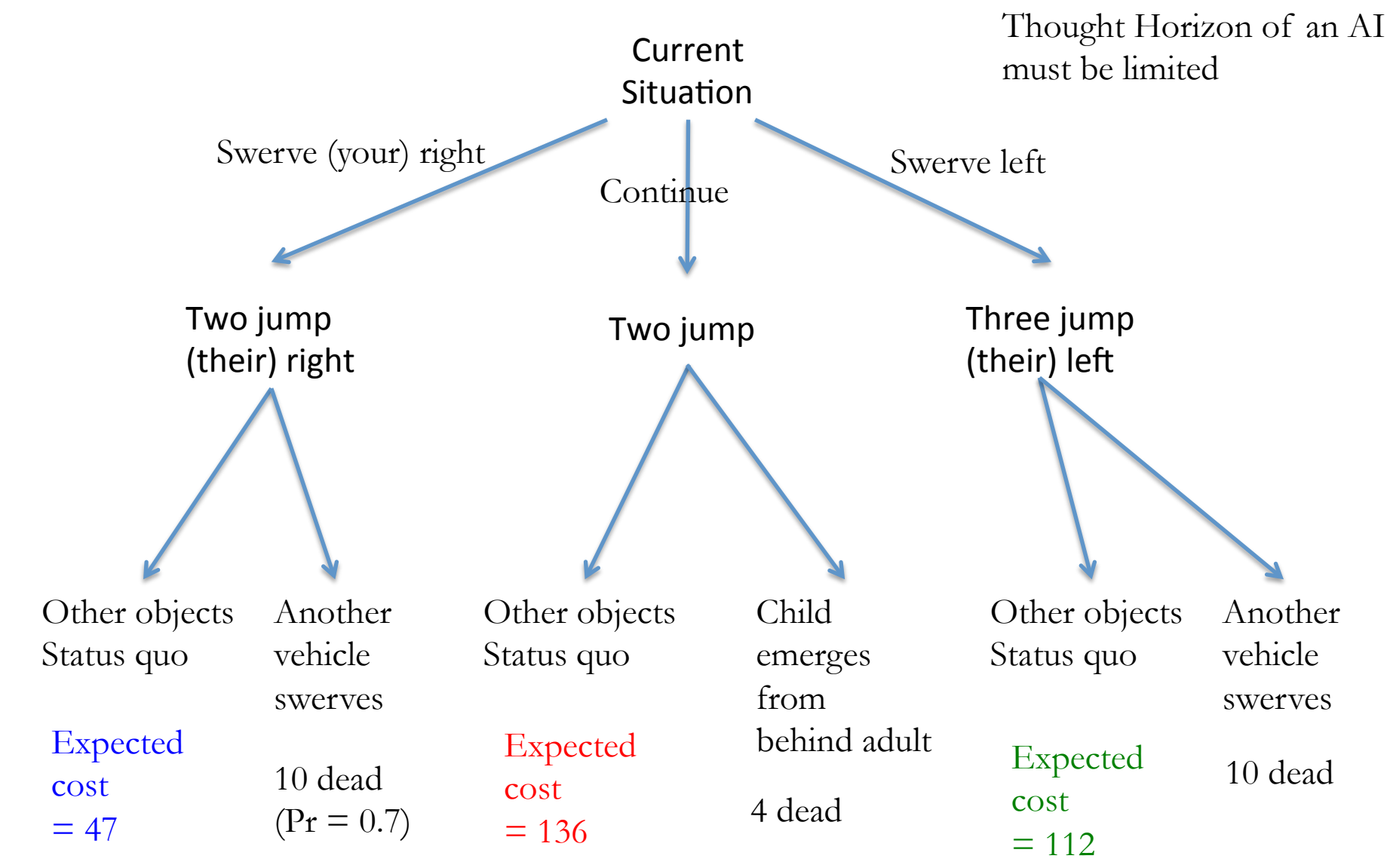
- The environment of the last slide is presumably stochastic, not determinsistic (see last week's Under the Hood) – why?

- Can you revise Asimov's Laws of Robotics be revised to take into account uncertainties (e.g., on probability of harm and expected amount of harm)?

- How are probabilities determined?
- In part, by expectation of what actions will be taken by other actors

- How are costs determined?
  - Is property damage relevant
  - Are animals relevant?

Lookahead (projecting into the future)



Communicating Among AIs

https://www.youtube.com/watch?v=4pbAI40dK0A

In a highly networked environment, vehicles won't be autonomous per se, but the will be "smart"