What is Acceptably Safe for Reinforcement Learning?

John Bragg and Ibrahim Habli

1st Workshop on Artificial Intelligence Safety Engineering (WAISE)
18th September, 2018
Västerås, Sweden

MBDA
MISSILE SYSTEMS

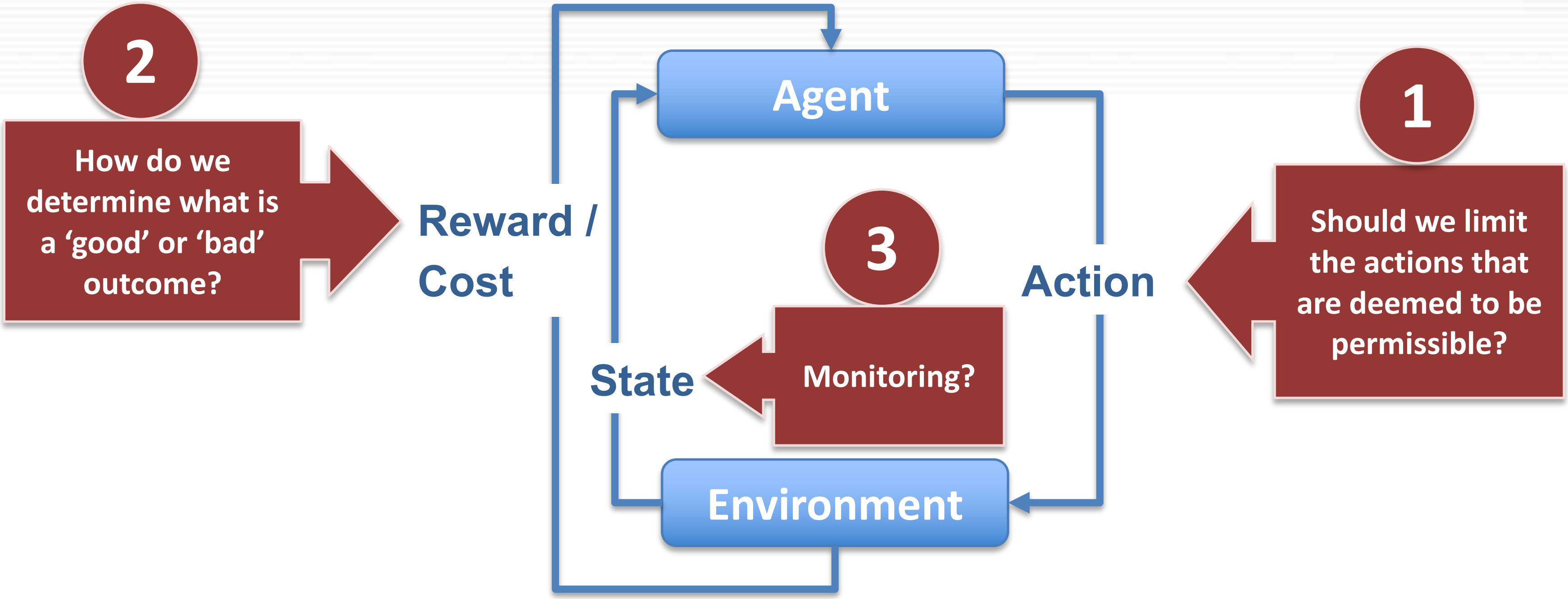
UNIVERSITY of York

Disclaimer

This presentation is an overview of academic research and is released for information purposes only.

The contents of this presentation and any views or opinions expressed are solely those of the authors and do not necessarily represent those of MBDA UK Ltd or the University of York.

A Simplified View of Reinforcement Learning


2
How do we determine what is a ‘good’ or ‘bad’ outcome?
Reward / Cost
Agent
3
Monitoring?
State
Environment
Action
1
Should we limit the actions that are deemed to be permissible?

1) Are we able to determine what the permissible actions are?
   - If we do this are we likely to unnecessarily constrain the agent in carrying out what actually might be the correct course of action.
2) ‘Reward’ vs. ‘Cost’ – the agent is going to learn both ‘good’ and ‘bad’ as a result of its actions, i.e. a negative outcome is as useful as a positive one.
3) Should we monitor the system and potentially ‘fail safe’ if there are indications that it is moving into an ‘unsafe’ state?

Safety of Complex Engineering Systems

- Safety-I : The ‘**Traditional**’ Approach
  - 👎 Focused largely on ‘predictable’ systems.
  - 👎 Approach to constructing a safety argument is based around assumptions/constraints at design-time.
  - 👎 Don’t really address software, which is the vehicle of implementation for ML/RL systems.
- Safety-II: The ‘**Adaptive**’ Approach
  - 👍 Focus is on the systems’ ability to adapt and succeed under varying conditions.
  - 👍 Safety analysis is predicated on the system’s ability to avoid hazardous conditions or deal with them when they occur.
  - 👎 Still to be demonstrated in ‘real-life’.

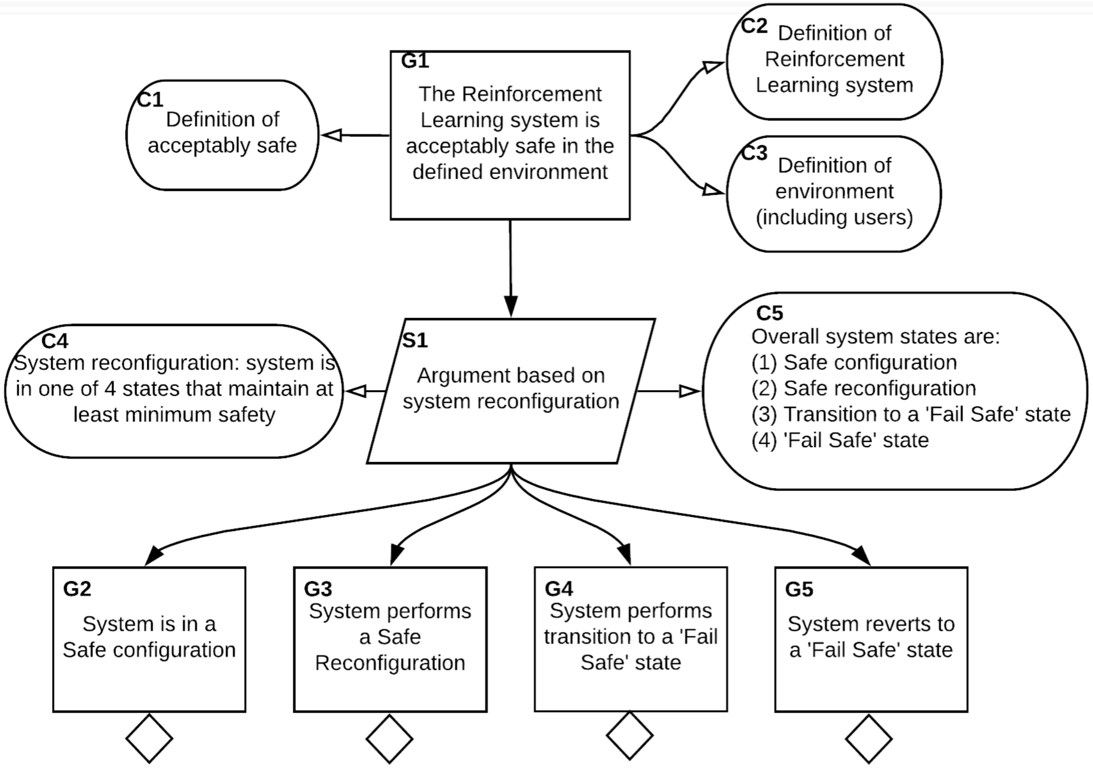
Reinforcement Learning and Safety

G2) System is in a Safe Configuration.

G3) System performs a Safe Reconfiguration.

G4) System transitions to a ‘Fail Safe’ state.

G5) System reverts to a ‘Fail Safe’ state.


G1
The Reinforcement Learning system is acceptably safe in the defined environment
C1
Definition of acceptably safe
C2
Definition of Reinforcement Learning system
C3
Definition of environment (including users)
S1
Argument based on system reconfiguration
C4
System reconfiguration: system is in one of 4 states that maintain at least minimum safety
C5
Overall system states are:
(1) Safe configuration
(2) Safe reconfiguration
(3) Transition to a 'Fail Safe' state
(4) 'Fail Safe' state
G2
System is in a Safe configuration
G3
System performs a Safe Reconfiguration
G4
System performs transition to a 'Fail Safe' state
G5
System reverts to a 'Fail Safe' state

Main strategy is to provide an ‘Argument based on Safe Reconfiguration’ (It does not introduce any new hazardous behaviour)

Elements of the Safety Argument

- Risk vs. Benefits
  - A wrong decision may not necessarily lead to harm
- Cost vs. Reward
  - ‘Overly-protective’ vs. ‘Overly-risky’
- Leading and Lagging Indicators
  - Identifying a move to a hazardous condition
- Safe Constraints on Learning
  - ‘Safe’ learning environments
- Fail Safe
  - Predicated on the RL Systems’ own ability to recognise it is no longer operating safely

Challenges Presented by Reinforcement Learning

Challenge Area	Short Description
Societal Acceptance	How do we choose which human and social values to embody within the system?
Risk vs. Benefits	Do the benefits of deploying an RL system outweigh the safety risks? Is this a sufficient threshold?
Cost vs. Reward	How does an RL System learn where a negative outcome still needs to be 'safe'?
Monitoring & Feedback	What mechanisms should be used for an RL System to decide what it needs to reconfigure?
Learning Constraints	Can safety only be achieved through constraining the way in which an RL System learns?
Fail Safe	How and when should RL Systems fail safe?
Intelligent Safety	How can an RL System update its own safety case and explain the choices it has made?

AI-based Safety Assurance and Explainability

- A human's ability to rationalise the choice that a machine has made is likely to be very limited.

- One possible solution, using Dynamic Safety Cases, would be to have the Safety Argument and monitoring evolve alongside the system.
- Explainable AI - MIT have taught a neural network how to show its own work.

Summary

- Constraining RL algorithms in the way in which they learn will help with the safety assurance argument, but this could both limit its functionality and possibly its ability to remain safe.

- Implementing ‘Intelligent Safety’, whereby the safety monitoring evolves alongside the system, could provide a solution, but only if the appropriate ‘cost’ and ‘reward’ mechanisms are selected.

- RL systems could be allowed to create and update its own safety case as it learns, but only if it can explain the rationale behind the decisions it has made.