Steps to Verifying Autonomous Systems

Michael Fisher

Department of Computer Science, University of Liverpool, UK

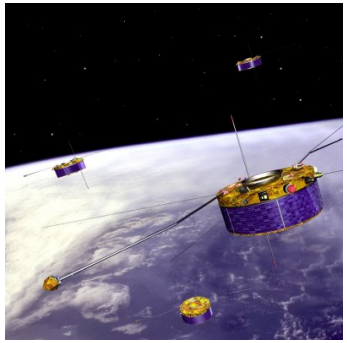
From July 2020: University of Manchester, UK

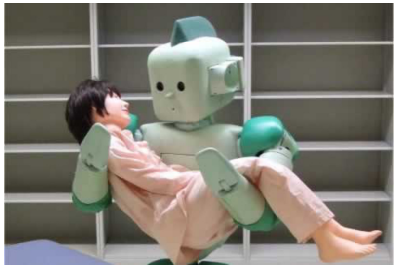
June 4, 2020

Autonomous Systems

Autonomy:

the ability of a system to make its own decisions and to act on its own, and to do both without direct human intervention.





rtc.nagoya.riken.jp/RI-MAN

www.volvo.com

Who makes the Decisions?

Even within 'autonomy', there are important variations concerning decision-making:

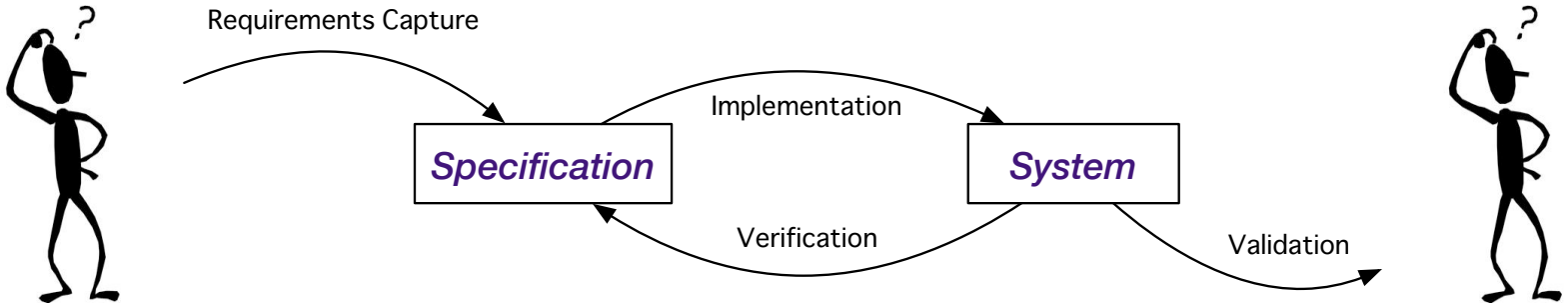
Automatic: involves a number of fixed, and prescribed, activities; there may be options, but these are generally fixed in advance.

Adaptive: improves its performance/activity based on feedback from environment — typically developed using tight continuous control and optimisation, e.g. feedback control system.

Autonomous: decisions made based on system's (belief about its) current situation at the time of the decision — environment still taken into account, but internal motivations/beliefs are important.

Distinguishing *between* these variations is often crucial.

Verification and Validation



Verification, typically

- formal verification
- simulation-based testing
- physical testing

Validation, typically

- physical testing
- user validation
- test scenarios

Verification = "are we building the system right?"
Validation = "are we building the right system?"

Testing *versus* Formal Verification

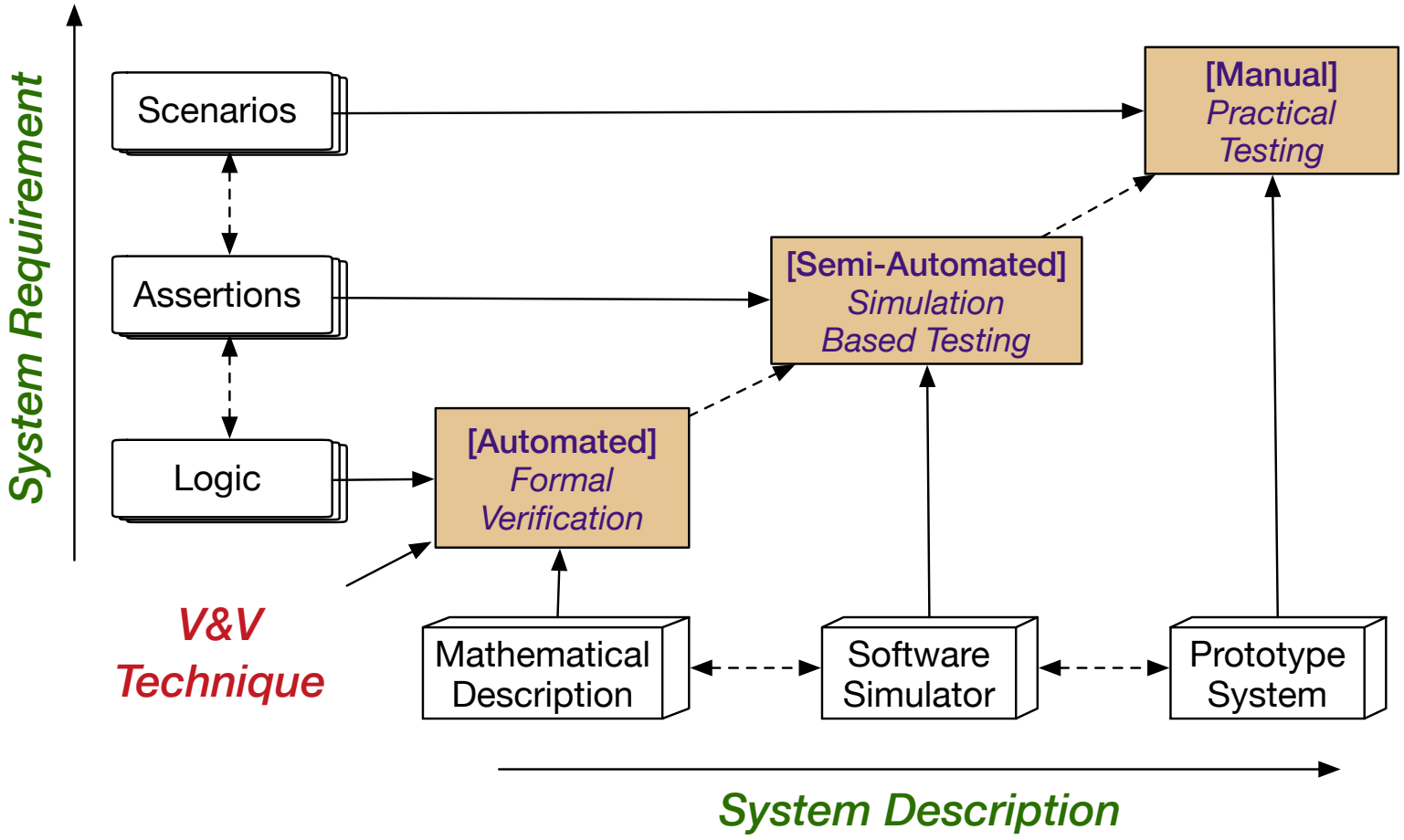
Advantages of Formal Verification:

- exhaustive — covers all situations;
- constructive — can ask what might cause a problem.

Disadvantages of Formal Verification:

- expensive — takes significant modelling/design effort;
- real-world — requires some abstraction/analysis.

Dimensions of Verification



[*A corroborative approach to verification and validation of human-robot teams*]

Verifying Decision-Making

Automatic: a number of fixed, and prescribed, activities.

Formally verify (fixed) behaviour.

Adaptive: e.g. feedback control systems.

Decisions effectively made through interaction with environment.

Can test or approximate (stochastic modelling) the system

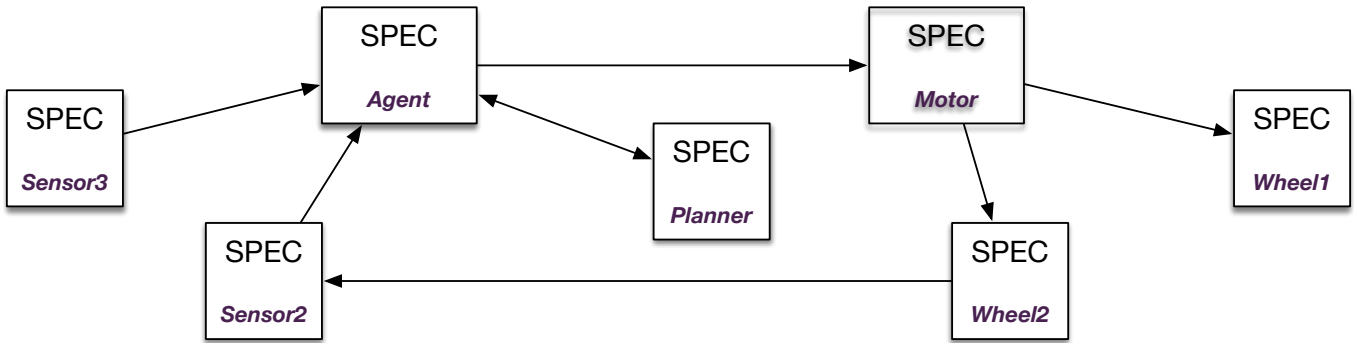
(but can never fully capture real environment)

Autonomous: internal decision-making process.

Formally verify the way the system makes its decisions

Robot Architectures to the Rescue

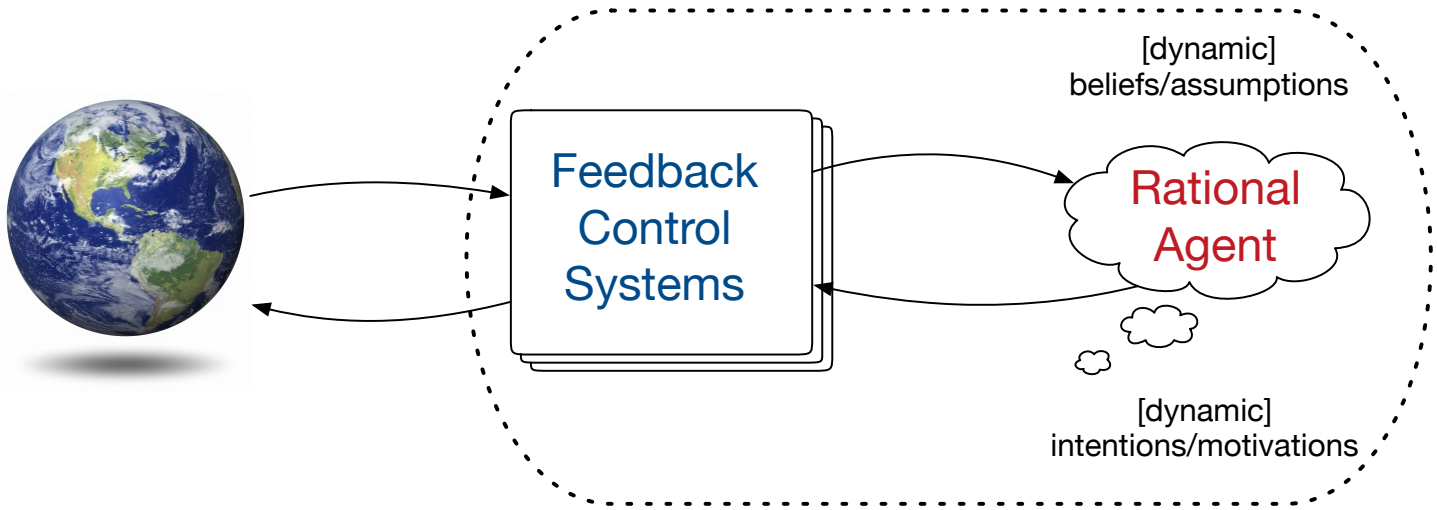
- Modularity — ROS, *ISO Modularity standard (22166)*, etc
- Transparency — *IEEE Transparency of Autonomous Systems standard (P7001)*, etc
- Verifiability — a concise, ideally *formal*, specification of the anticipated/expected behaviour of each module:



Our Approach

Our approach is that

we should be certain what the autonomous system **intends** *to do and how it* **chooses** *to go about this*

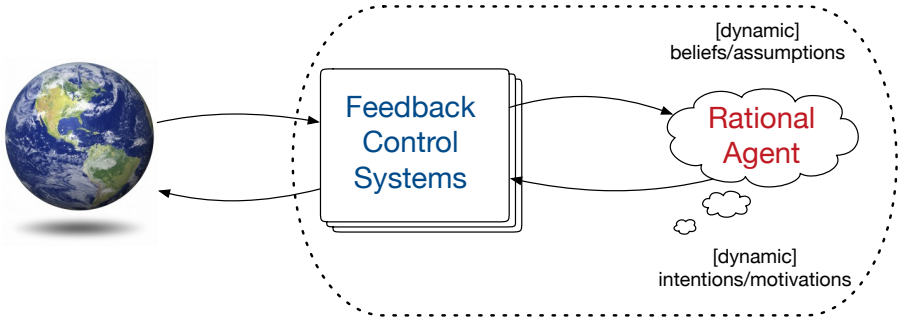


A *rational agent* (typically, a BDI Agent):

must have explicit **reasons** *for making the choices it does, and should expose/explain these when needed*

Heterogeneous Approaches

We can employ different verification techniques to different modules/components.



- We can *formally verify* the rational agent's decision-making
  $\longrightarrow$ we can be certain about this.
- We can *simulate/test* or *verify/monitor* the feedback control
- We can *practically test* whole system
  $\longrightarrow$ provides some practical confidence?

Reliable Autonomy

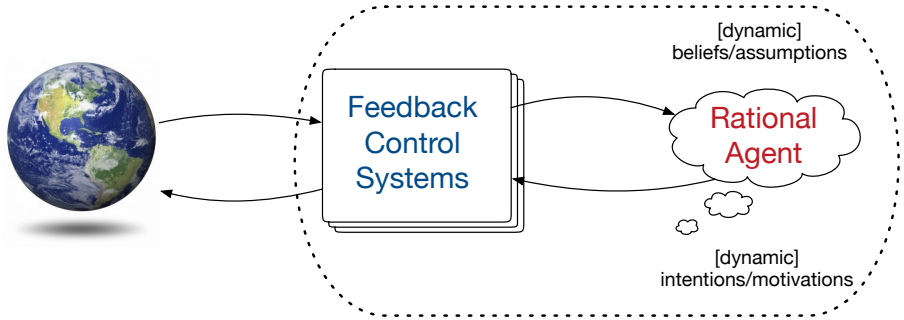
Especially with autonomous robotics, in safety-critical scenarios,

it is vital that we use strong (e.g. formal) verification techniques for any component/software we rely on

As a corollary,

never fully **rely** on any component that has not been formally verified

With increasing autonomy, we rely on the decision-making agent, and so must be **sure** it will do what we want.



Sample Relevant Publications

- Bremner, Dennis, Fisher, Winfield. On Proactive, Transparent, and Verifiable Ethical Reasoning for Robots. *Proceedings of the IEEE*, 2019.
- Cardoso, Farrell, Luckcuck, Ferrando, Fisher. Heterogeneous Verification of an Autonomous Curiosity Rover. In *NASA Formal Methods Symposium*, 2020.
- Cardoso, Ferrando, Dennis, Fisher. An Interface for Programming Verifiable Autonomous Agents in ROS. In *European Conference on Multi-Agent Systems*, 2020.
- Dennis, Fisher. Verifiable Self-Aware Agent-Based Autonomous Systems. *Proceedings of the IEEE*, 2020
- Dennis, Fisher, Lincoln, Lisitsa, Veres. Practical Verification of Decision-Making in Agent-Based Autonomous Systems. *Journal of Automated Software Engineering*, 2016.
- Farrell, Cardoso, Dennis, Dixon, Fisher, Kourtis, Lisitsa, Luckcuck. Modular Verification of Autonomous Space Robotics. In *Assurance of Autonomy for Robotic Space Missions Workshop*, 2019.
- Ferrando, Dennis, Ancona, Fisher, Mascardi. Recognising Assumption Violations in Autonomous Systems Verification. In *International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, 2018.
- Kamali, Dennis, McAree, Fisher, Veres. Formal Verification of Autonomous Vehicle Platooning. *Science of Computer Programming*, 2017.
- Luckcuck, Farrell, Dennis, Dixon, Fisher. Formal Specification and Verification of Autonomous Robotic Systems: A Survey. *ACM Computer Surveys*, 2019.
- Fisher, Dennis, Webster. Verifying Autonomous Systems. *Communications of the ACM*, 2013
- Webster, Cameron, Fisher, Jump. Generating Certification Evidence for Autonomous Unmanned Aircraft Using Model Checking and Simulation. *Journal of Aerospace Information Systems*, 2014.
- Webster, Western, Araiza-Illan, Dixon, Eder, Fisher, Pipe. A Corroborative Approach to Verification and Validation of Human–Robot Teams. *International Journal of Robotics Research*, 2020.