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.
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?”
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.
[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:

```
SPEC Motor
SPEC Agent
SPEC Sensor3
SPEC Sensor2
SPEC Planner
SPEC Wheel2
SPEC Wheel1
SPEC Planner
```

Michael Fisher
Steps to Verifying Autonomous Systems
Our Approach

Our approach is that

\textit{we should be certain what the autonomous system intends to do and how it chooses to go about this}

A \textit{rational agent} (typically, a BDI Agent):

\textit{must have explicit reasons for making the choices it does, and should expose/explain these when needed}
We can employ different verification techniques to different modules/components.

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


