

Architectural Concepts for Human-Rated Automation

David Wagner

California Institute of Technology

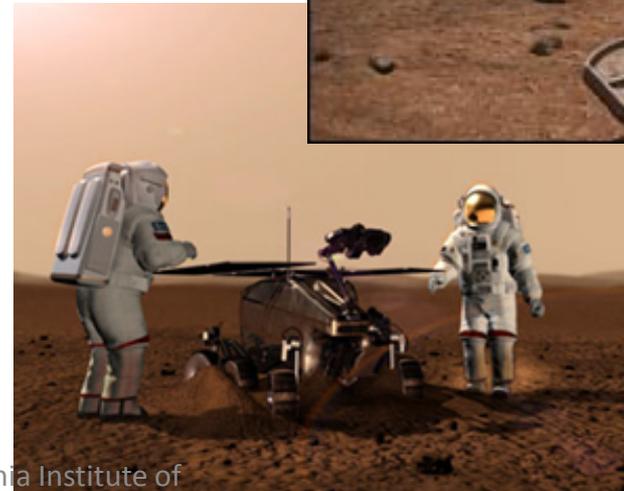
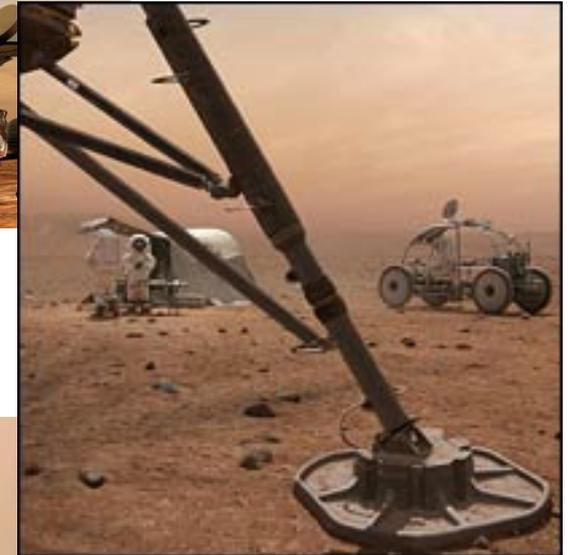
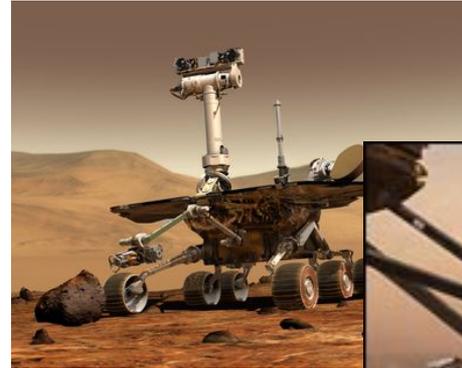
Jet Propulsion Laboratory

Outline

- The need for autonomy (1p)
- Architectural Qualities (1p)
- Goal-Oriented Control
- Safeguards
- Enhanced situational awareness

The Need for Automation

- Need for autonomy
 - Communications latency and bandwidth
 - Crew must be able to respond to situations on their own
- Time limit for a safe response
- Complexity of response coordination in time and space
- Cost-effective operations



System Safety Through Architecture

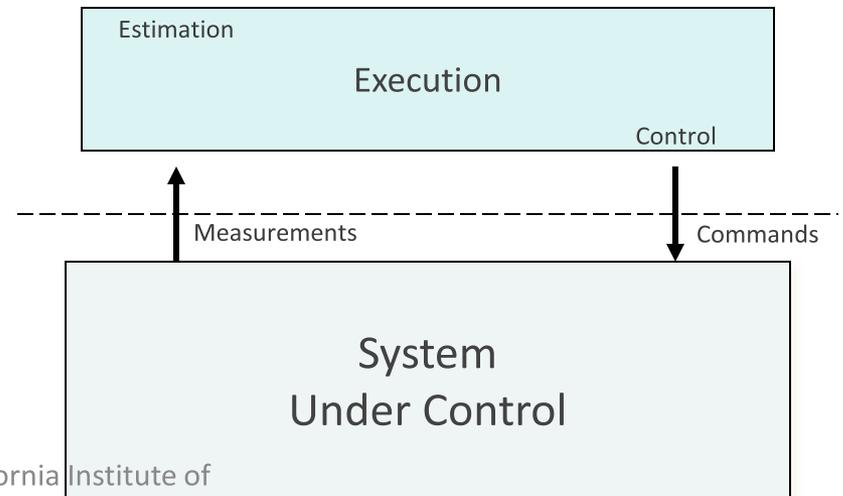
- Desired Qualities
 - Transparency
 - Predictability
 - Dependability
 - Fault tolerant reliability
 - Verifiability
 - Safety Cognizance

Recommendation: Spend more time up front in requirements analysis and architecture to *really* understand the job and its solution

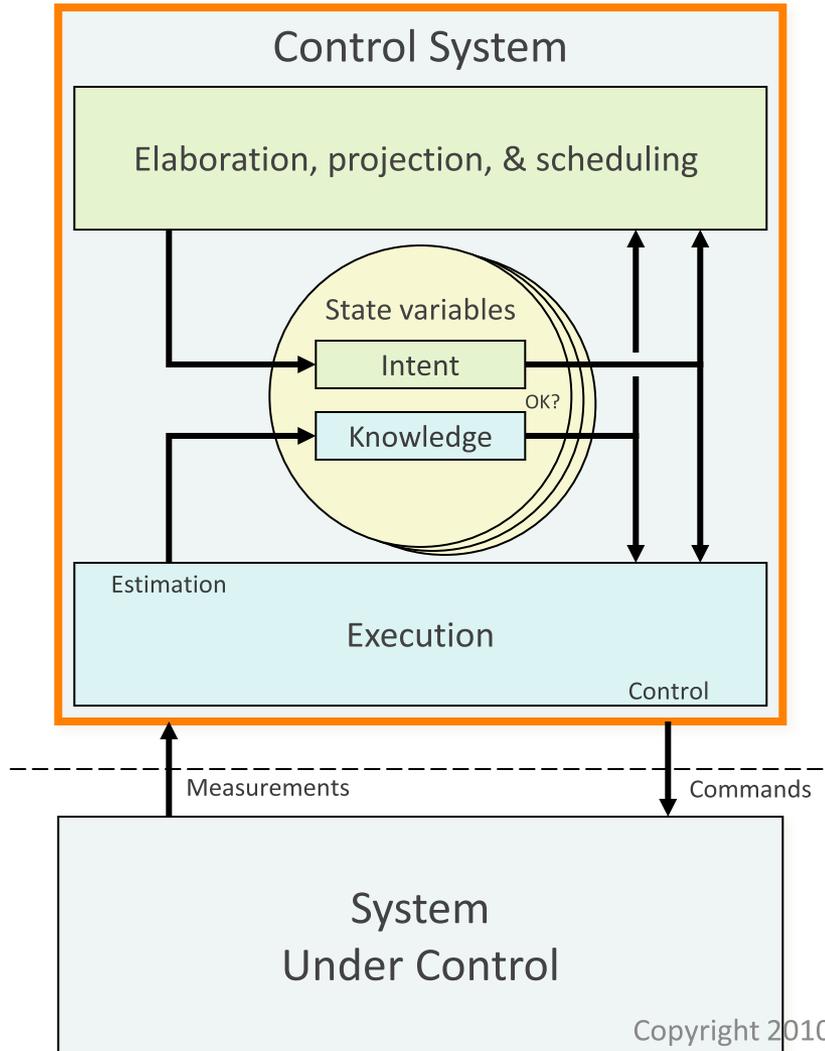
[NASA Study on Flight Software Complexity, NASA Office of Chief Engineer, 2009](#)

Traditional Control

- Immediate commands
- External coordination
 - Ground-based
 - Manual (procedure)
 - Manually planned
- Scripting
- Command Sequencing
- Sequential, Imperative programming



Goal-Oriented Control

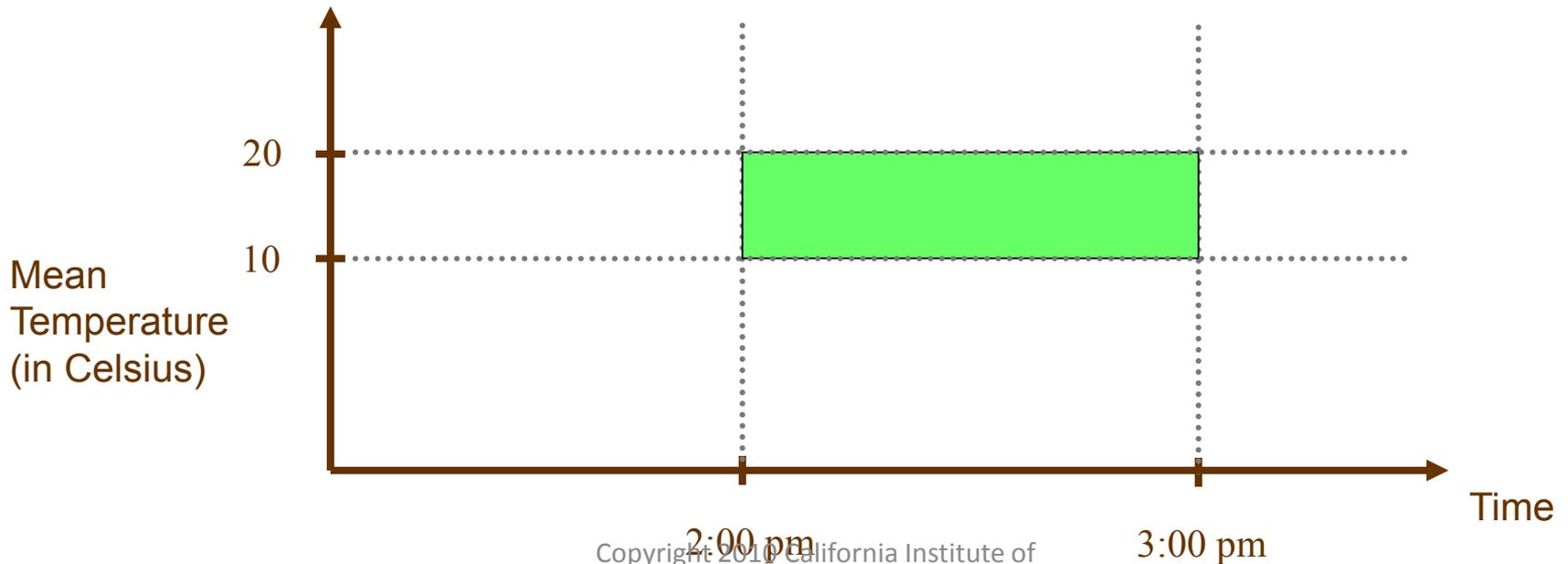


- Goals express intent
 - Explicit constraints on state over time
 - Elaborate supporting goals
- Scheduling layer
 - Verifies achievability prior to execution
- Execution layer executes verified plan
 - Can react to faults by replanning

A Goal in State Analysis

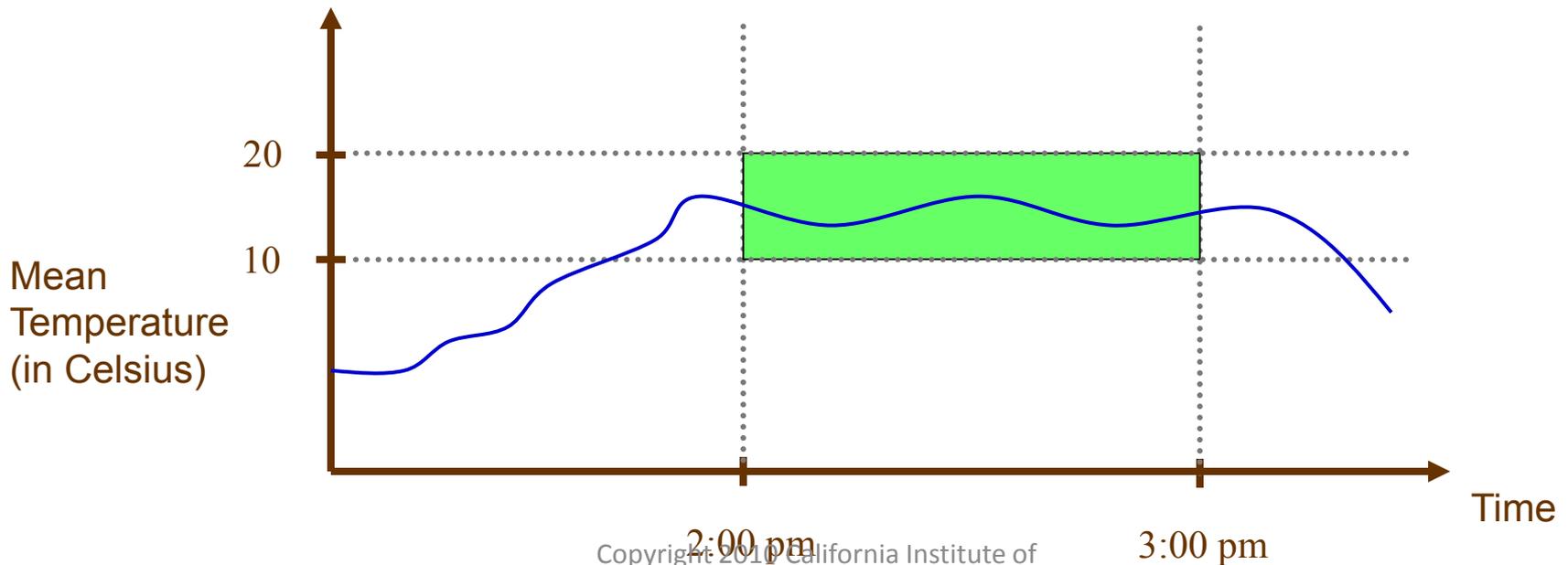
- Goals are conditions that persist over some time interval, and provide a statement of operational intent...

“Camera mean temperature is in the range 10–20°C from 2:00 pm to 3:00 pm”



A Goal that Succeeded During Execution

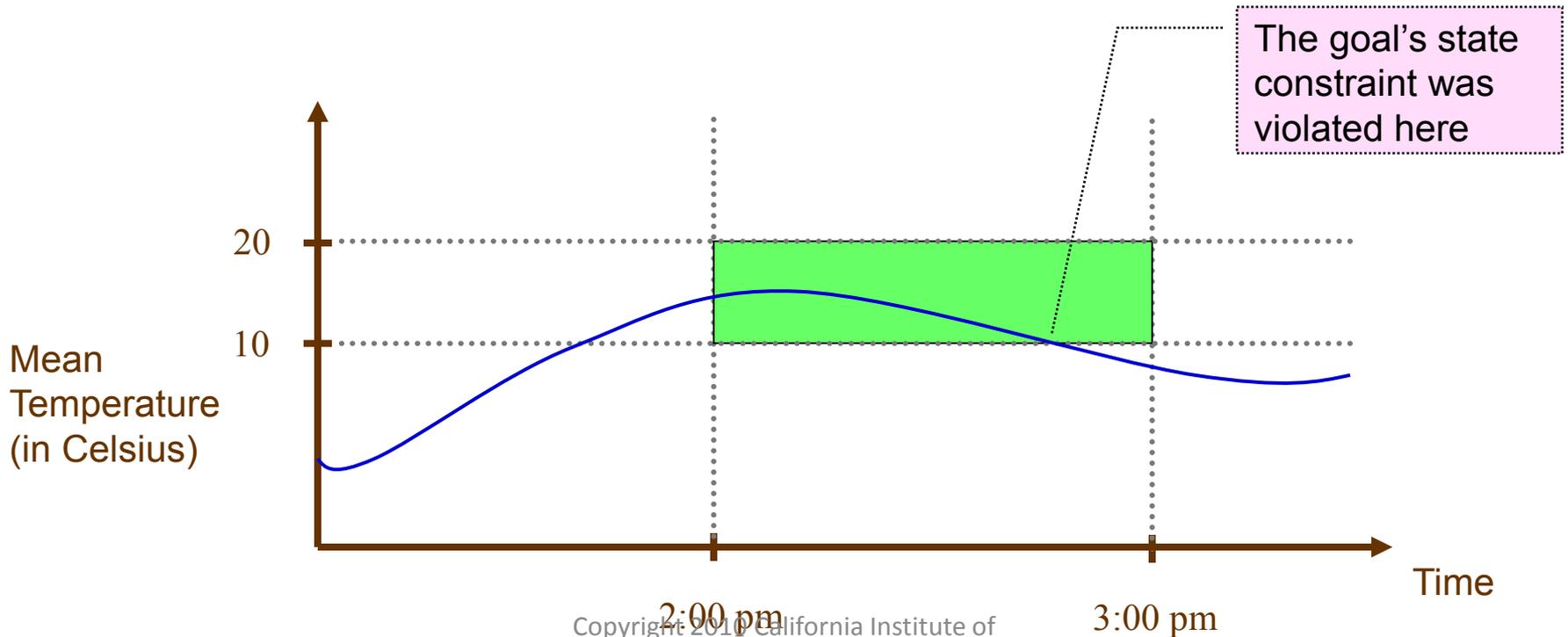
- This state history satisfies the goal
 - Mean value in blue
- “Camera mean temperature is in the range 10–20°C from 2:00 pm to 3:00 pm”



A Goal that Failed During Execution

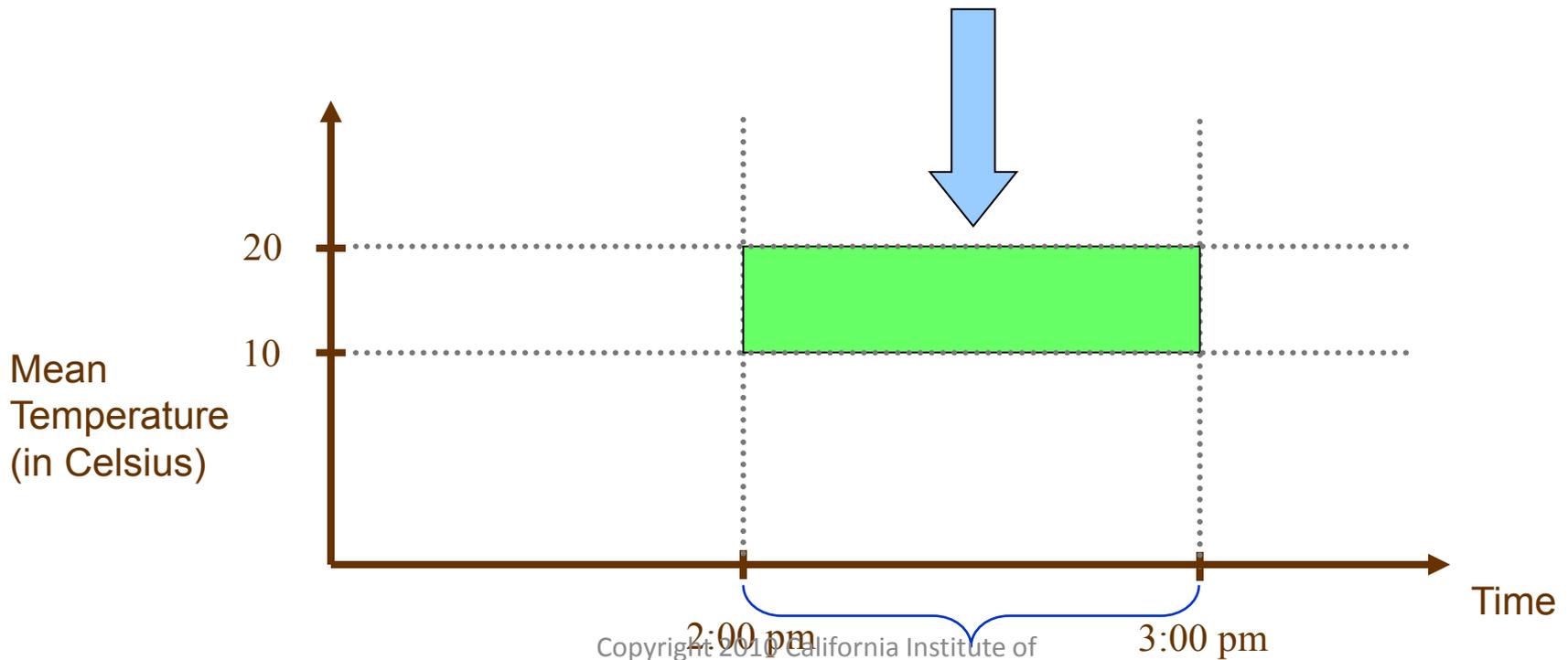
- This state history does *not* satisfy the goal
- Mean value in blue

“Camera mean temperature is in the range 10–20°C from 2:00 pm to 3:00 pm”



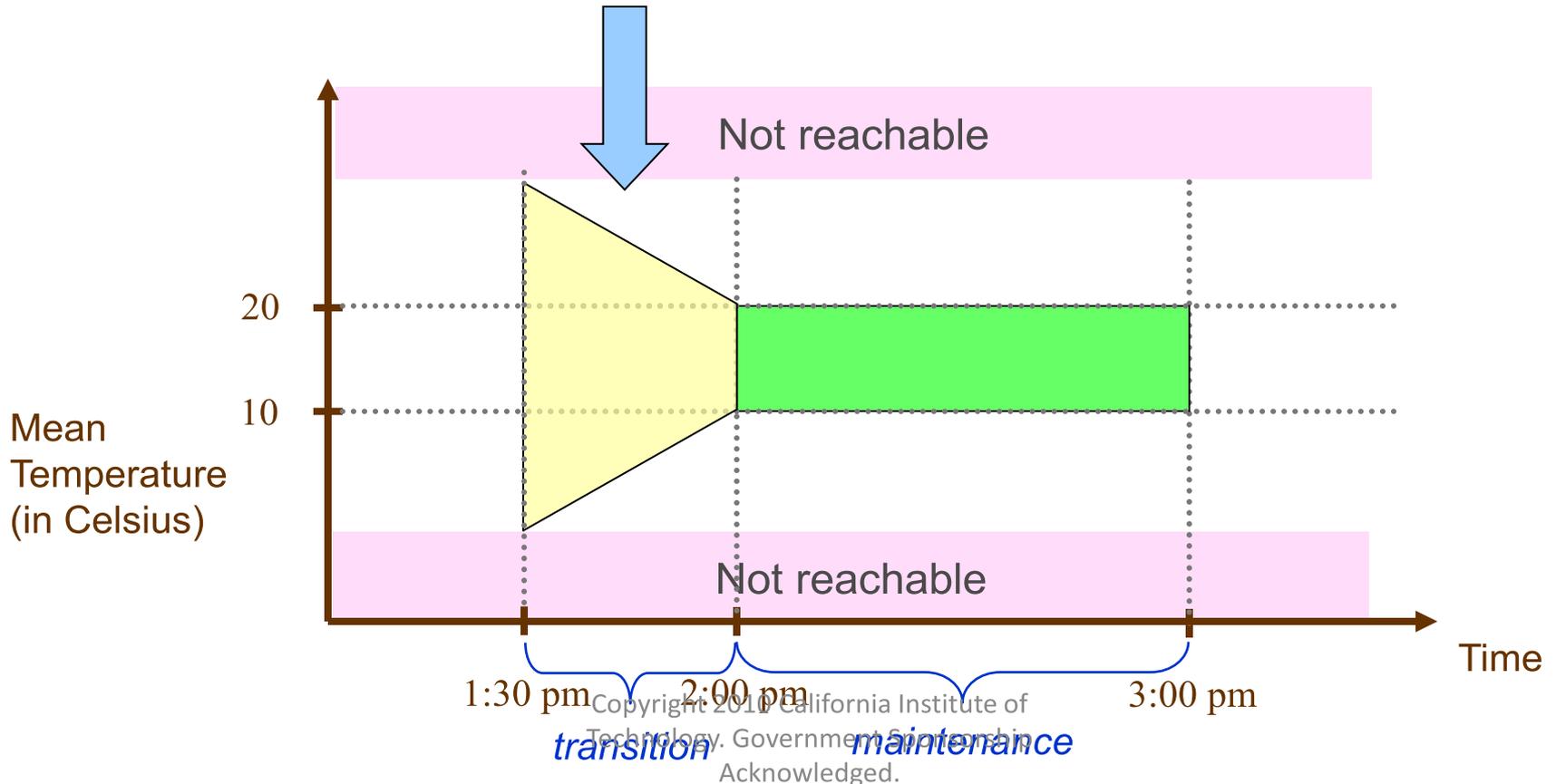
Maintenance Goals

- The intent is to maintain temperature in a range, but
 - This goal will fail unless temperature already in range at 2:00 pm
- “Camera mean temperature is in the range 10–20°C from 2:00 pm to 3:00 pm”



Transition Goals

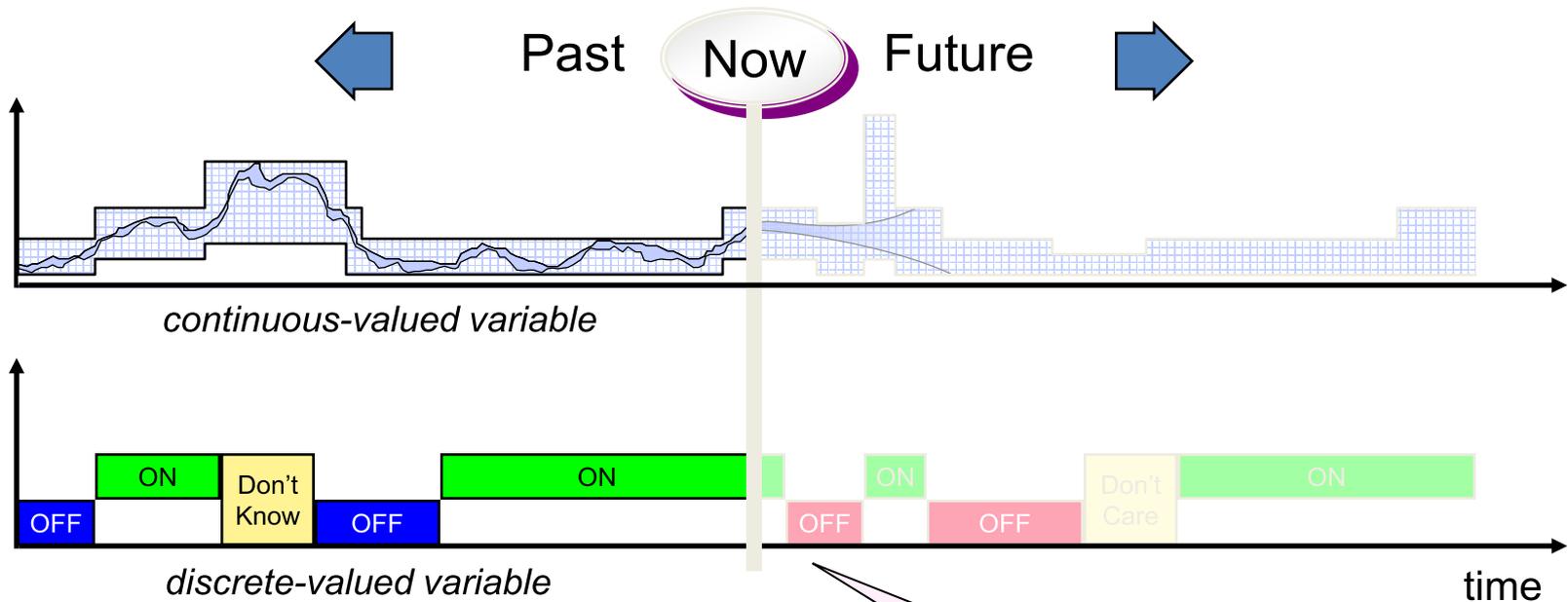
- A transition goal often precedes a maintenance goal
 - This goal gets temperature in range by 2:00 pm
- “Camera mean temperature is in the range 10–20°C by 2:00 pm”



The Time Line Concept

History compared to plans

Predictions informed by plans



Goal Achievement

- Goals generally have an intended state change or transition of some controllable state to coordinate
 - “Drive to waypoint”
- Goals can also describe flight rules on any observable state
 - “maintain at least 15% battery margin”
 - “do not point camera at the sun”
- Goal violation can trigger a re-planning response

Safeguards

- Ordinary Goals have a specified outcome
 - Open a valve
 - Take a picture
 - Drive to waypoint
- Only need to specify one tactic/method to achieve it
- Goals can also specify flight rules, or safety limits
 - Keep temperature within safe range
 - Don't point camera at the sun
 - Operate robot only when people aren't present

Safeguards

- Ordinary goals specify one or more methods or tactics describing a path to achievement
 - Intent is to control system into the desired state
- Safeguards don't care (or specify) how
 - Intent is to be able to react when/if constraint is violated
 - Reaction may include
 - Warning
 - Active avoidance
 - Safing
 - Replanning

Safeguard Examples

Enhanced Situational Awareness

- Progress of Execution
 - Success or failure of specific goals (“pre-breathing complete”)
 - Starting or ending a specific goal (“starting to depressurize”)
 - Awaiting human approval to proceed (“Ok to begin depress”)
 - Reminder that a goal is ongoing (“calibration is still in progress”)
- Changes in State
 - A subsystem diagnosed as unhealthy has degraded (“UHF radio failed diagnostic”)
 - Critical resource quantity reported in units of time (“30 minutes of suit oxygen remaining”)
 - Long-term trend in a state variable’s value (“CO2 concentration is rising”)
 - Unusual values or patterns in a state value history (“voltage is fluctuating”)
 - Proximity warning: threshold is exceeded (“battery is below 25%”)
 - Discrepancy between human-provided input and evidence from other sources
- Timing
 - Estimation of when a goal will start (“egress will begin in 50 minutes”)
 - Amount of time left to complete a goal (“battery recharge complete in 40 minutes”)
- Control Authority
 - Change in control authority (“crew has control of robot”)

Managed Control Authority

- Management through access controls
 - Can at warn of *potential for* conflict
 - Can't deal with long-running or scheduled control activities
- Goal-Oriented Coordination
 - Goals have to merge into a single coordinated plan
 - New goals have to merge into executing plan or be rejected



Tunable Automation

- All the low-level command and measurement interfaces are still there
- Safeguard constraints can be individually added/removed
- Goals can depend on human interaction, execution
- Goal elaboration is scalable, composable
 - Combine two goals into a composite plan
 - Decompose a plan and only automate some subgoals
- Dependencies are explicit and transparent

Verifiability

- Two parts:
 - Planning/Execution Engine – well defined and tested behaviors
 - Goals, Goal compositions, Plans – flexible runtime definition
- Explicit intent enables
 - Formal verification
 - Runtime verification
- Mechanism can operate in an advisory mode where it only suggests control inputs to users, and easily extended to full control authority if found to be trustworthy

Conclusion

- Architectural approach enables efficient, flexible, and verifiable solution
- Safety and Operability are enhanced by
 - Plan verification
 - Continuous constraint execution
- Good systems and software architecture is the best defense against incidental complexity
 - “Point of view is worth 80 IQ points”
- Architecture is about principles
 - E.g., All control decisions should be based on estimated state, desired state, and models of behavior.

[NASA Study on Flight Software Complexity, NASA Office of Chief Engineer, 2009](#)