ESAW

26^{th} September 2008

Controlling the Global Behaviour of a Reactive MAS : Reinforcement Learning Tools

François Klein, Christine Bourjot, Vincent Chevrier
francois.klein@loria.fr

LORIA
Nancy Université
France

Outline

- Scientific context and issues
  - MAS and control
- Proposition of a dynamical solution
  - Using reinforcement learning tools
- Case study and assessment
  - On a toy example modelling pedestrians
- Conclusion and future works

Reactive multi-agent system

- Simple individual behaviours
  - System's dynamics defined at this local level
- Complex collective (emergent) behaviour
  - Observed at global level

- How to make the MAS show a particular (target) global behaviour ?

Issues in controlling a MAS

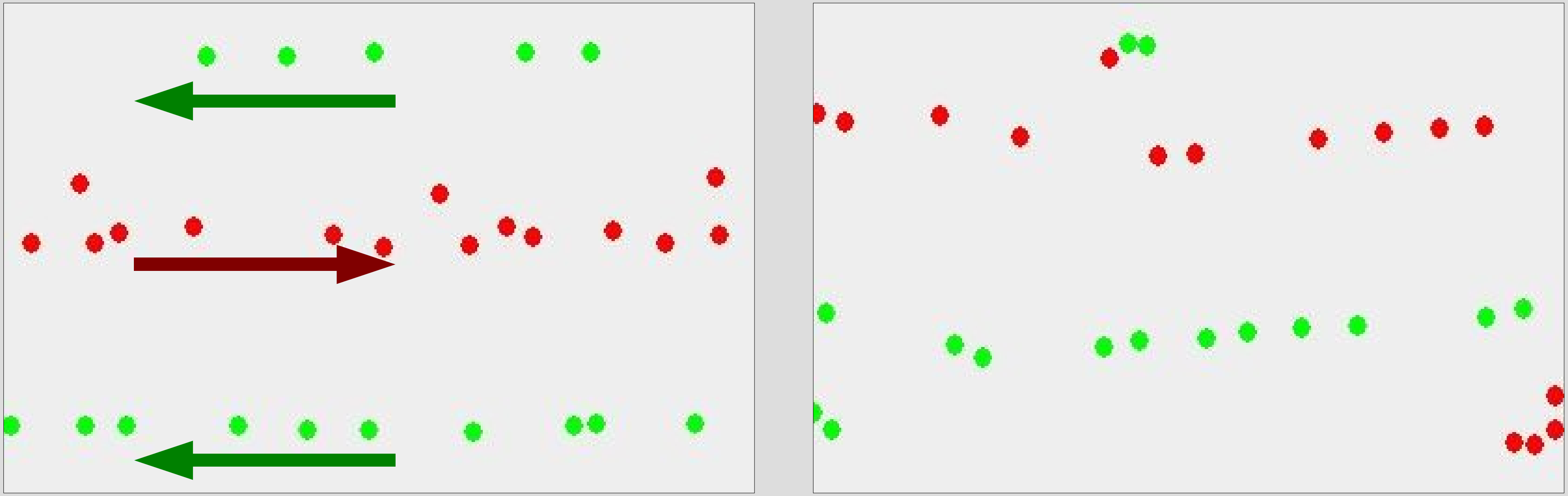
- The target stands at the **global level**
- The possible actions only affect the system's dynamics at **local level**

- Issues
  - Difficult to understand the local-global link
  - Strongly non-linear dynamics
  - The accurate consequences of an action are unpredictable
- **But** ∃ global regularities...

$\rightarrow$ Illustration on a toy example

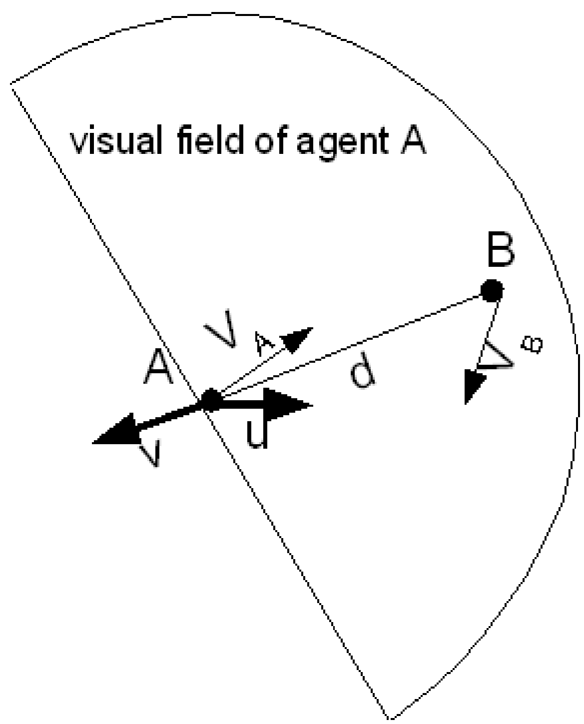
Toy example

- Agents : inspired by pedestrians
- Environment : torric corridor
- Emergent structures : lines and blocks



Toy example: agents' behaviour

- Forces-based behaviour
- 5 parameters



Movement force : $\vec{f}_M = C_M . \vec{u}$

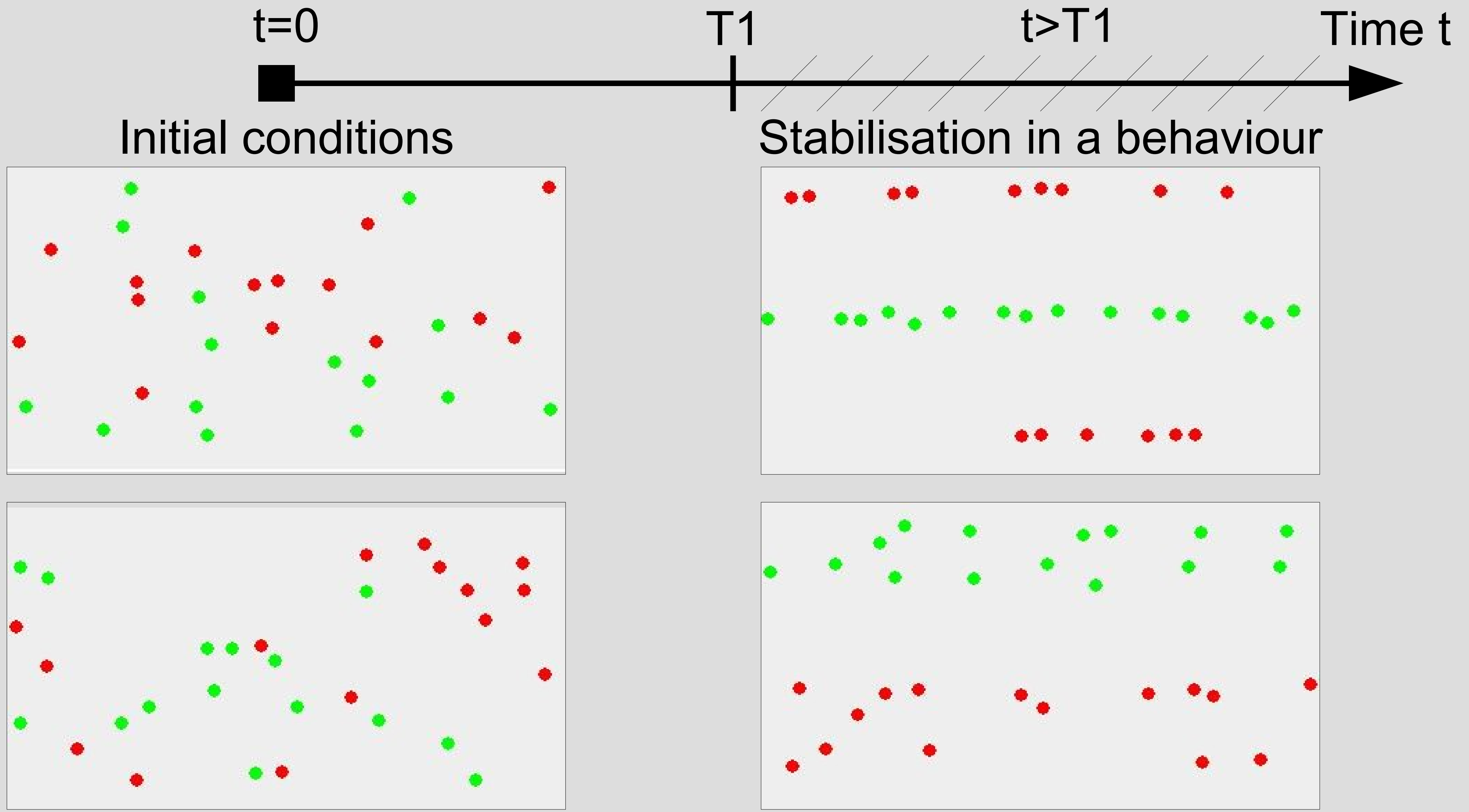
Separation force : $\vec{f}_S = \frac{C_S}{d}\vec{v}$

Avoidance force : $\vec{f}_A = C_A \frac{\vec{v} \times (\vec{v}_A - \vec{v}_B)}{d}\vec{v}$

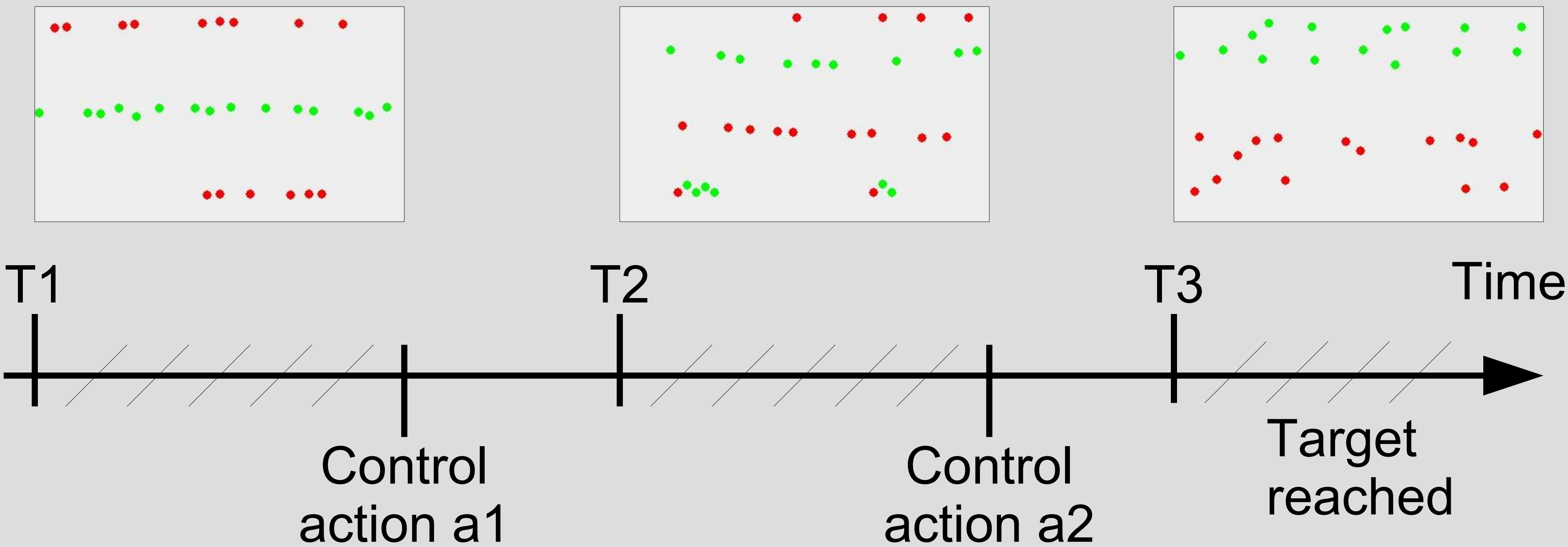
$$\vec{v}_A(t+1) = \vec{v}_A(t) + \sum \vec{f}$$

$$|\vec{v}_A(t+1)| = min\{|\vec{v}_A(t+1)|, M_S\}$$

Toy example: collective behaviour



Control of the pedestrians system



e.g. Change of the environment size

e.g. Change of the maximum speed

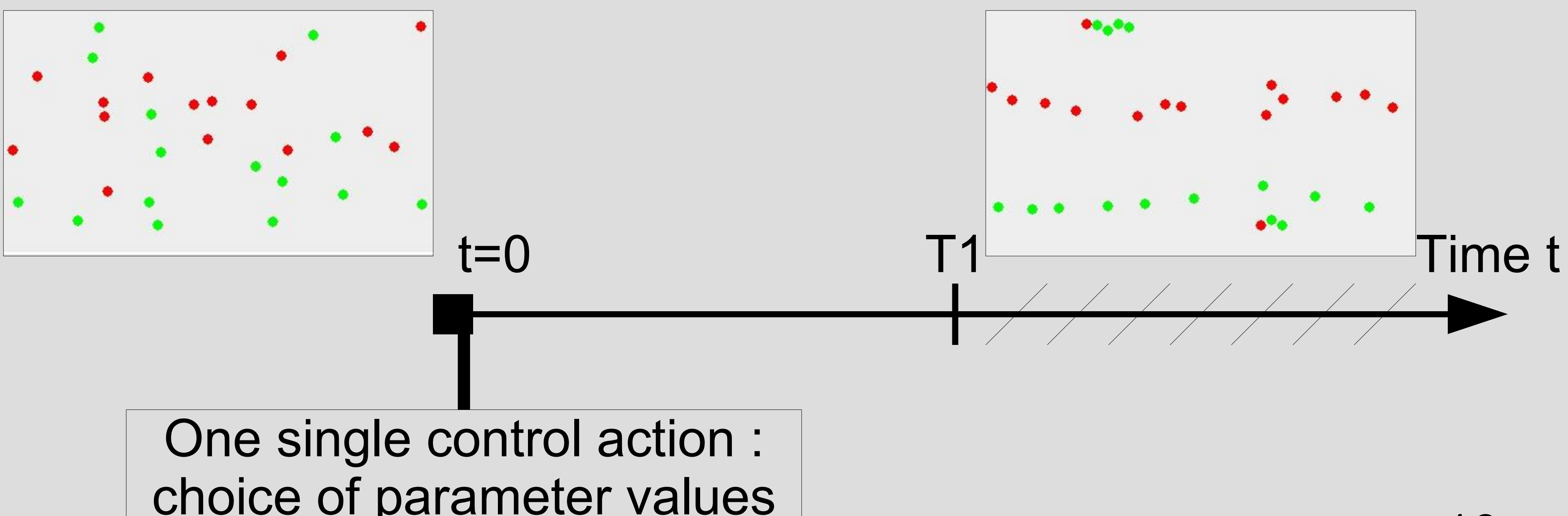
→ How to reach the target ?

How to control a MAS ?

- Analytical approach
  - Namely (global) differential equations
  - Unsufficient
    Wegner 1997, Edmonds 2004, DeWolf 2005
- Experimental approaches
  - Static (off-line)
  - Dynamical (on-line)

Static approaches

- (Sau 01), (DWo 05), (Feh 06), (Cal 05), (Bru 03)
- Engineering of the system
- Namely parameter setting
- Reduction of the experimental exploration



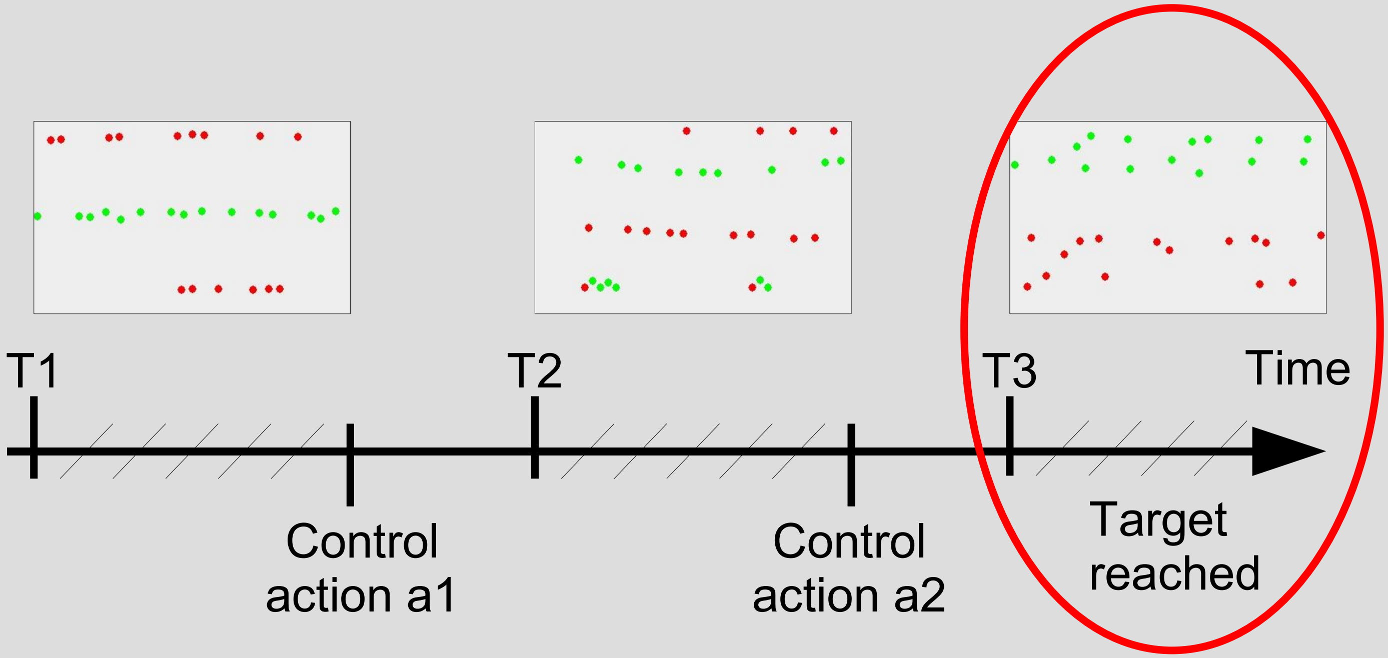
Dynamical approaches

- Heuristic global consideration
  - (Cam 04), (Ber 07)
  - No automatisation/optimisation in the choice of the actions
- Markov model approaches
  - (Tho 04), (Sut 98)
  - DEC-MDP (def. of the individual behaviours)
  - Usual application does not answer the control problem (action means, observation)
  - Complexity (Ber 02)

Proposition of a dynamical solution using RL tools

- Global behaviour determination

measurement

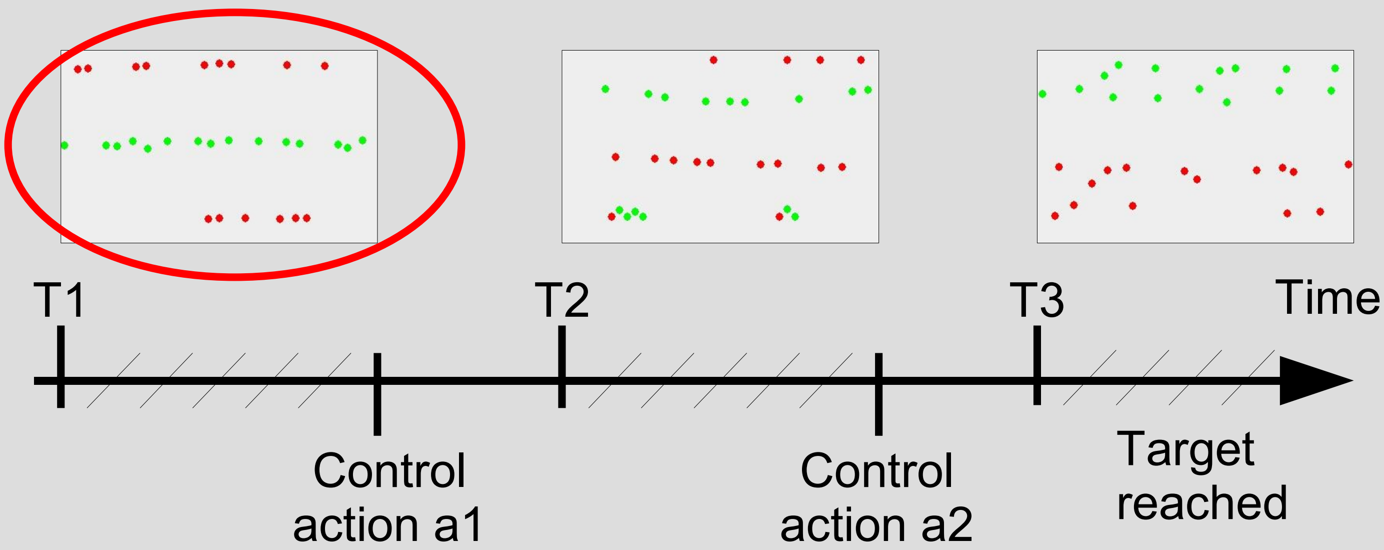


Proposition of a dynamical solution using RL tools

- Global behaviour determination
- Decision context

measurement

s



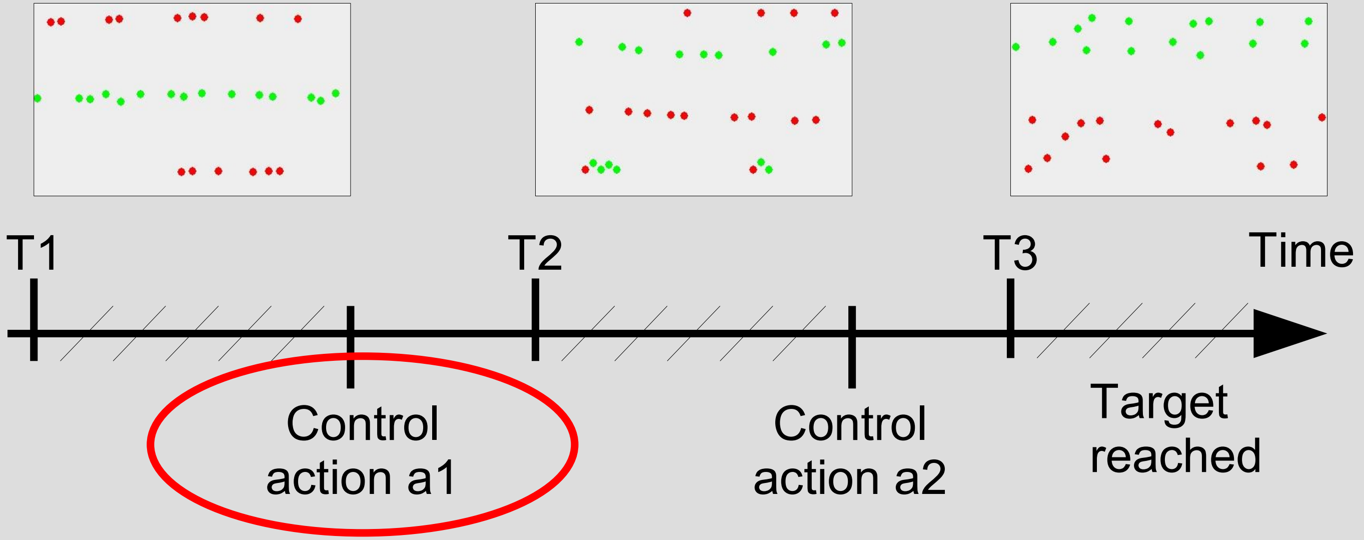
Proposition of a dynamical solution using RL tools

- Global behaviour determination
- Decision context
- Possible kinds of control actions

measurement

S

A



Proposition of a dynamical solution using RL tools

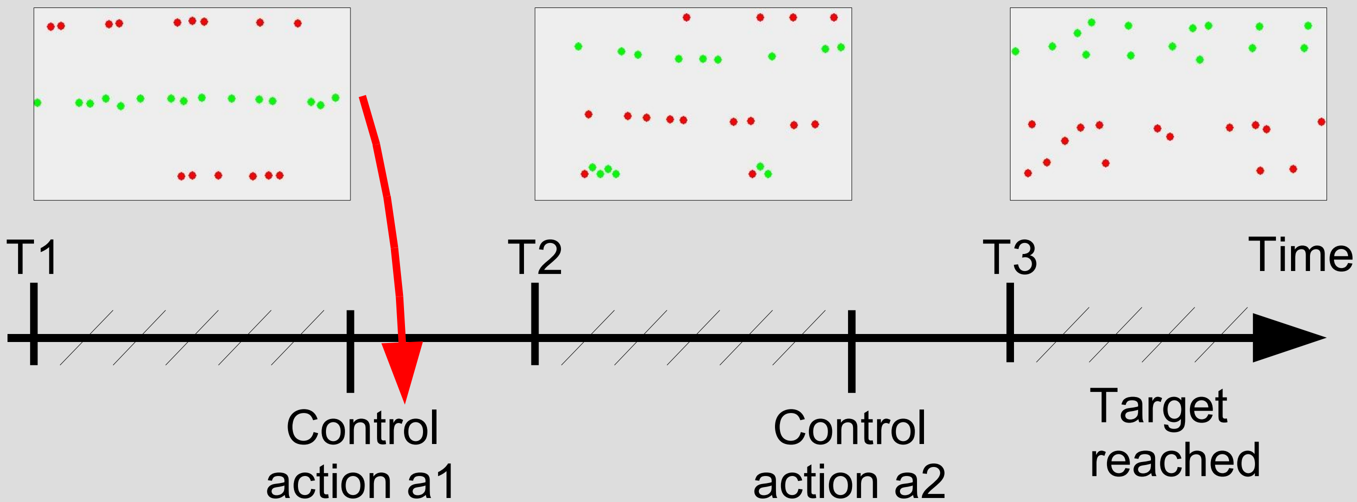
- Global behaviour determination
- Decision context
- Possible kinds of control actions
- Control action decision

measurement

S

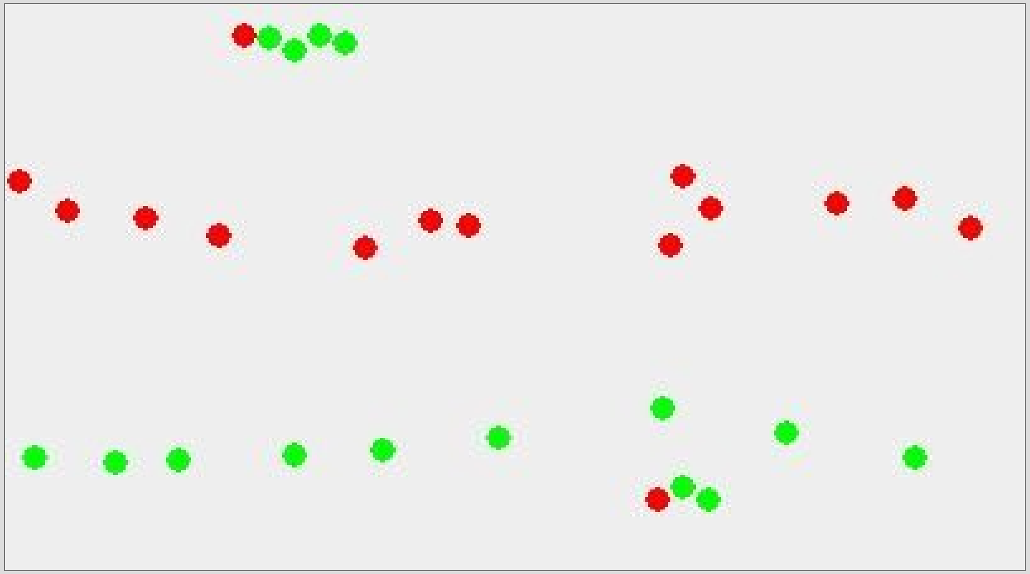
A

policy



Global behaviour determination

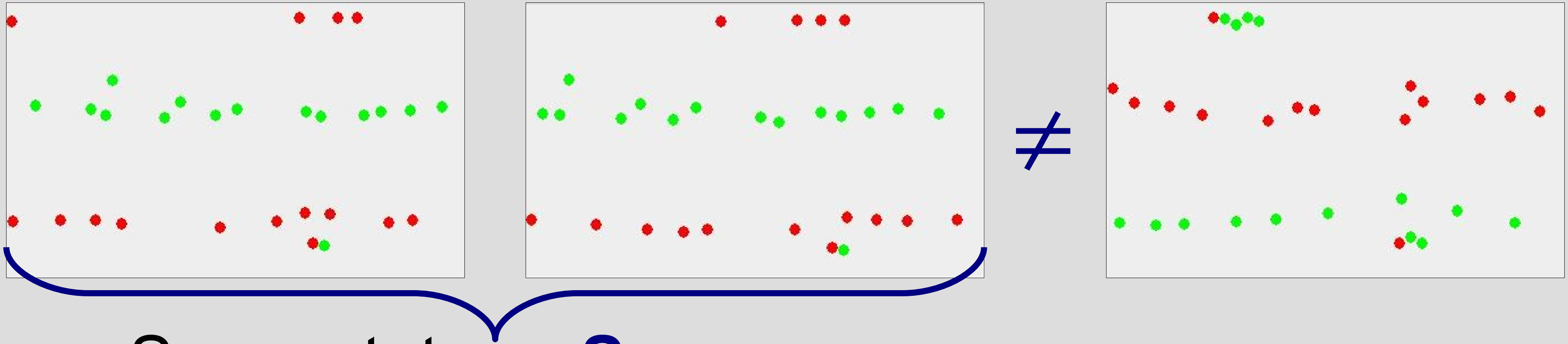
- Automatic global behaviour measurement
  - Formal characterisation of the target ≠ intuitive
  - Experimental → automatic method **measurement**



- Target = 2 lines **OK**
- Target = No blocks **NO**

Decision context

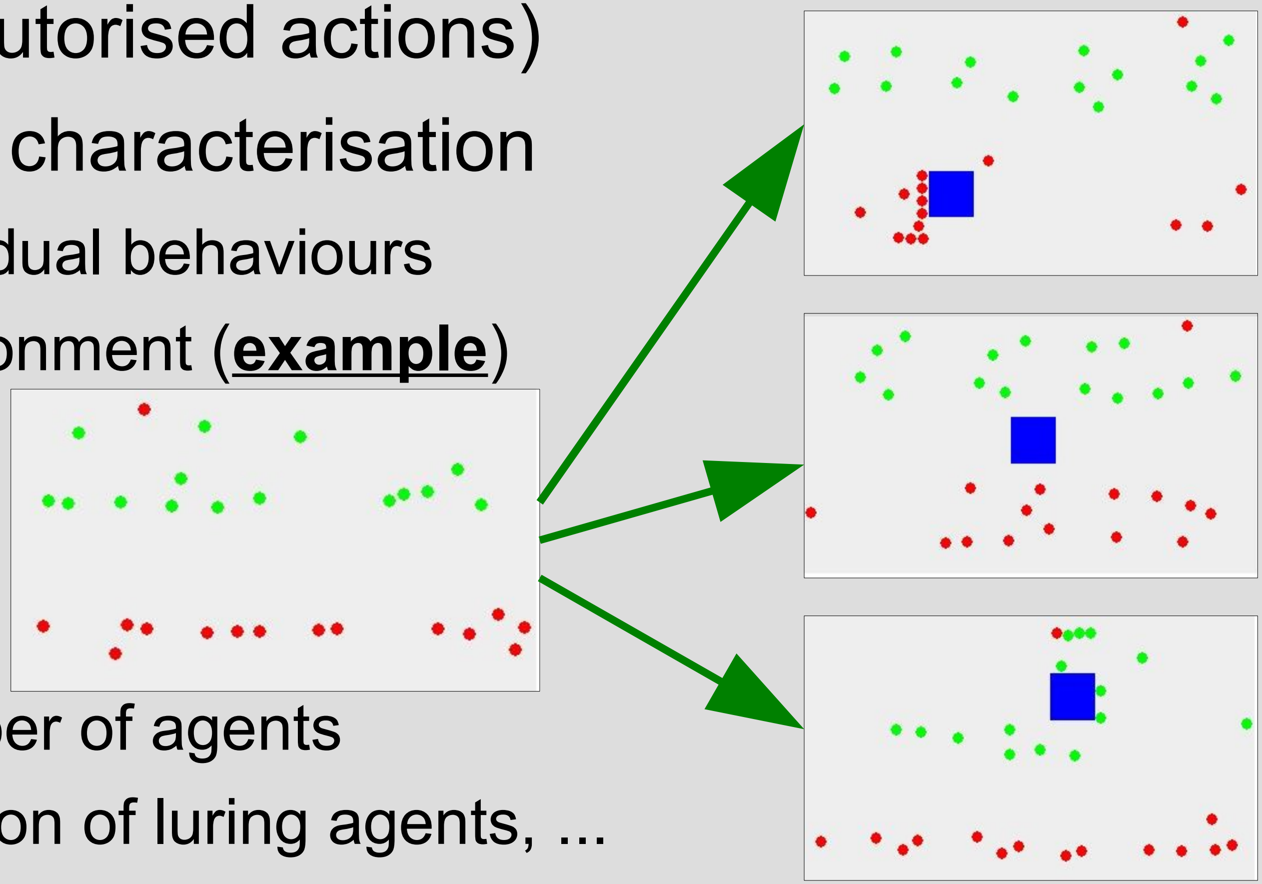
- Dynamical approach ⇒ distinction of situations
  - Differenciation of states **S**
  - Good choice (states level)
    - Few states = simpler = knowledge generalisation
    - Many states = more adequate actions



Same state s∈ **S**

Possible kinds of control actions

- Set **A** of possible actions
  - The controller can choose an action in **A** in each state (autorised actions)
  - Actions characterisation
    - Individual behaviours
    - Environment (**example**)



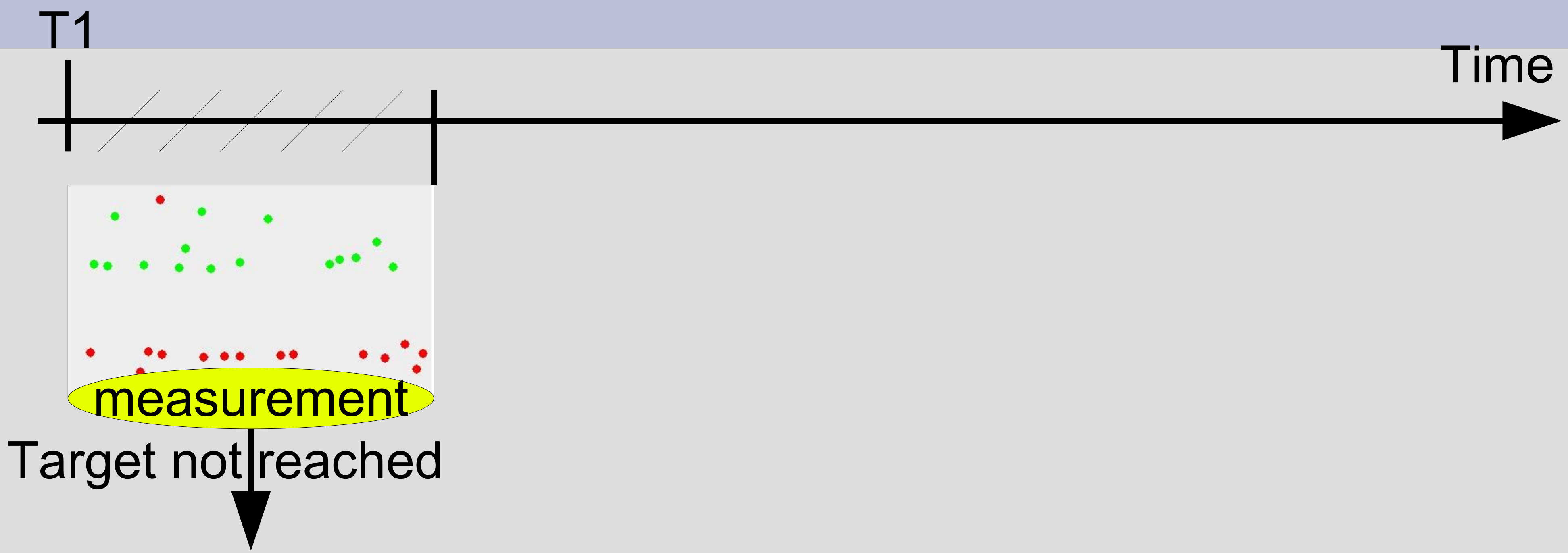
    - Number of agents
    - Addition of luring agents, ...

Control action decision

- Policy : function **S**→**A** to reach the target
- Computation
  - Use of reinforcement learning tools
  - Principle
    - A reward is granted to the tested actions if the target is reached → best actions in each state
  - Complexity reduction
    - Dynamic programming
    - Rationnal exploration: in each state, the more promising actions have their estimation refined

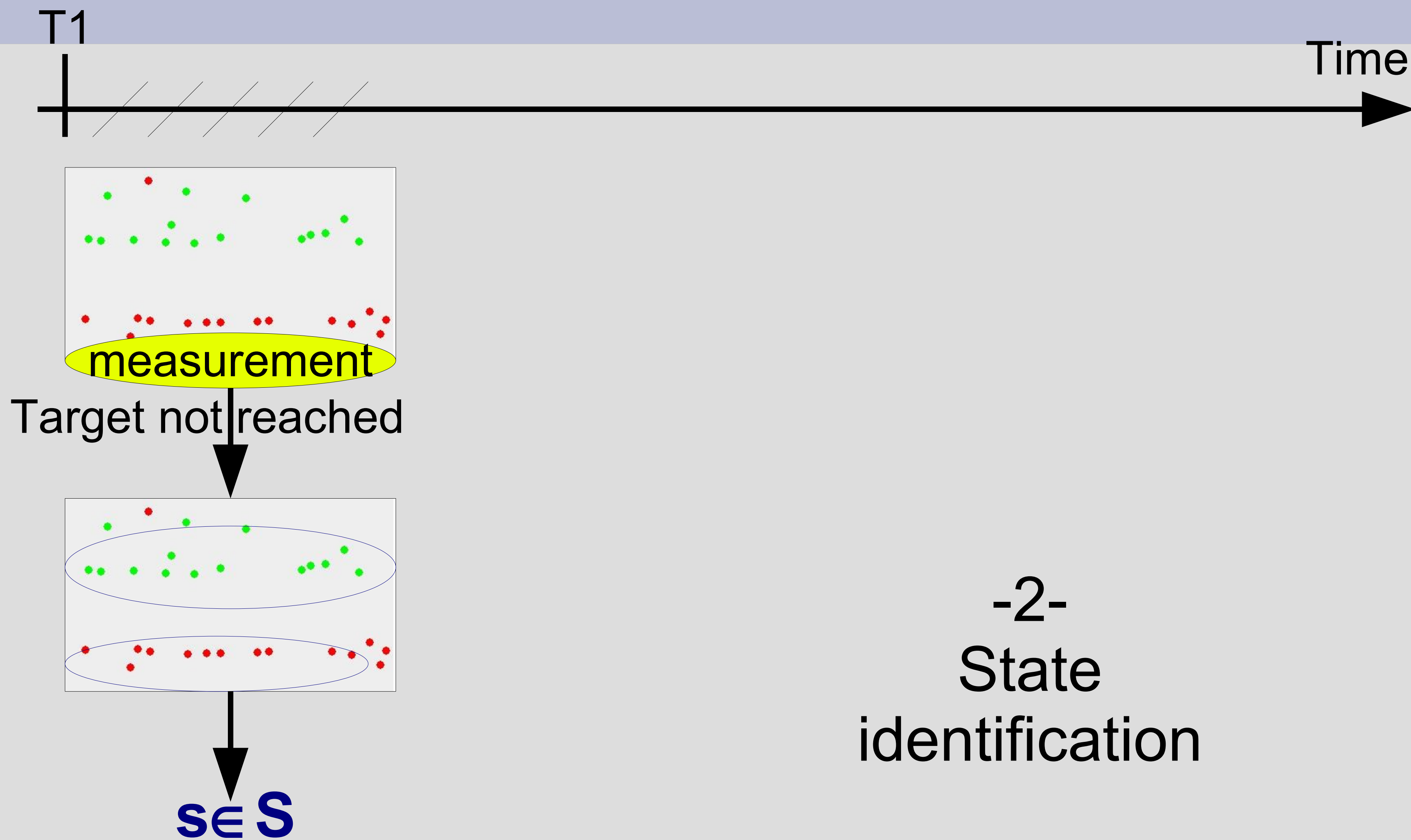
policy

Summary



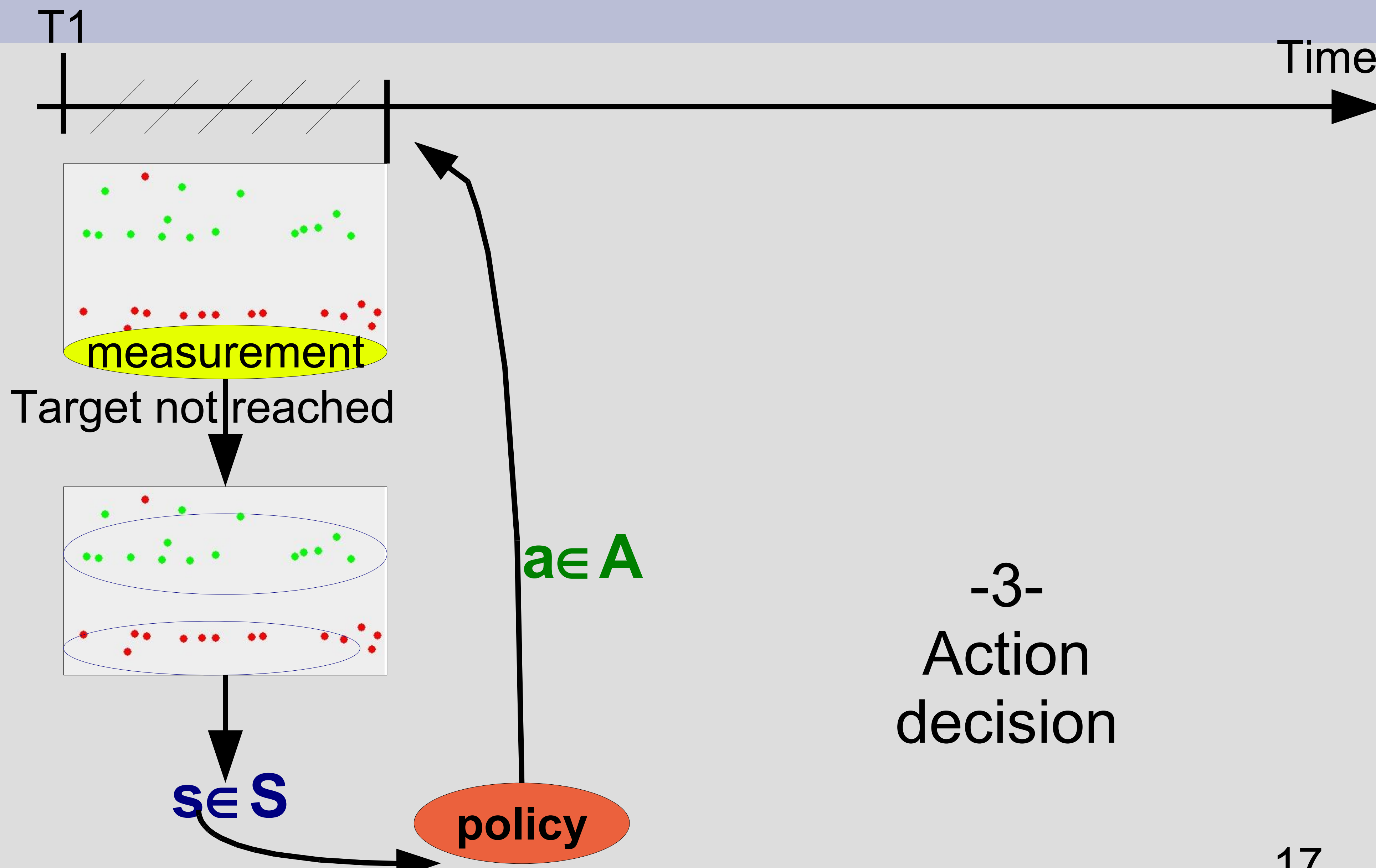
-1-
Behaviour
determination

Summary

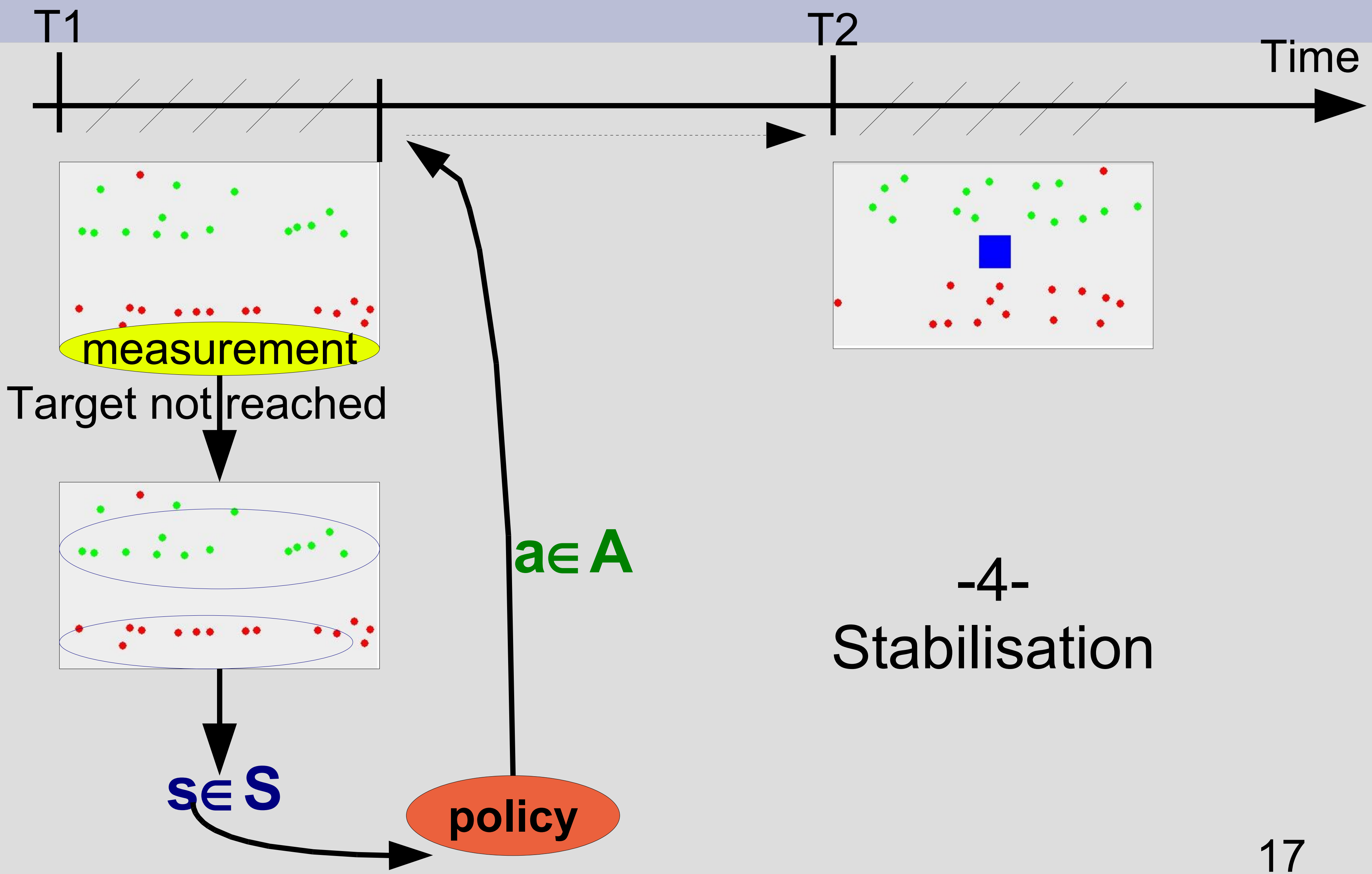


-2-
State
identification

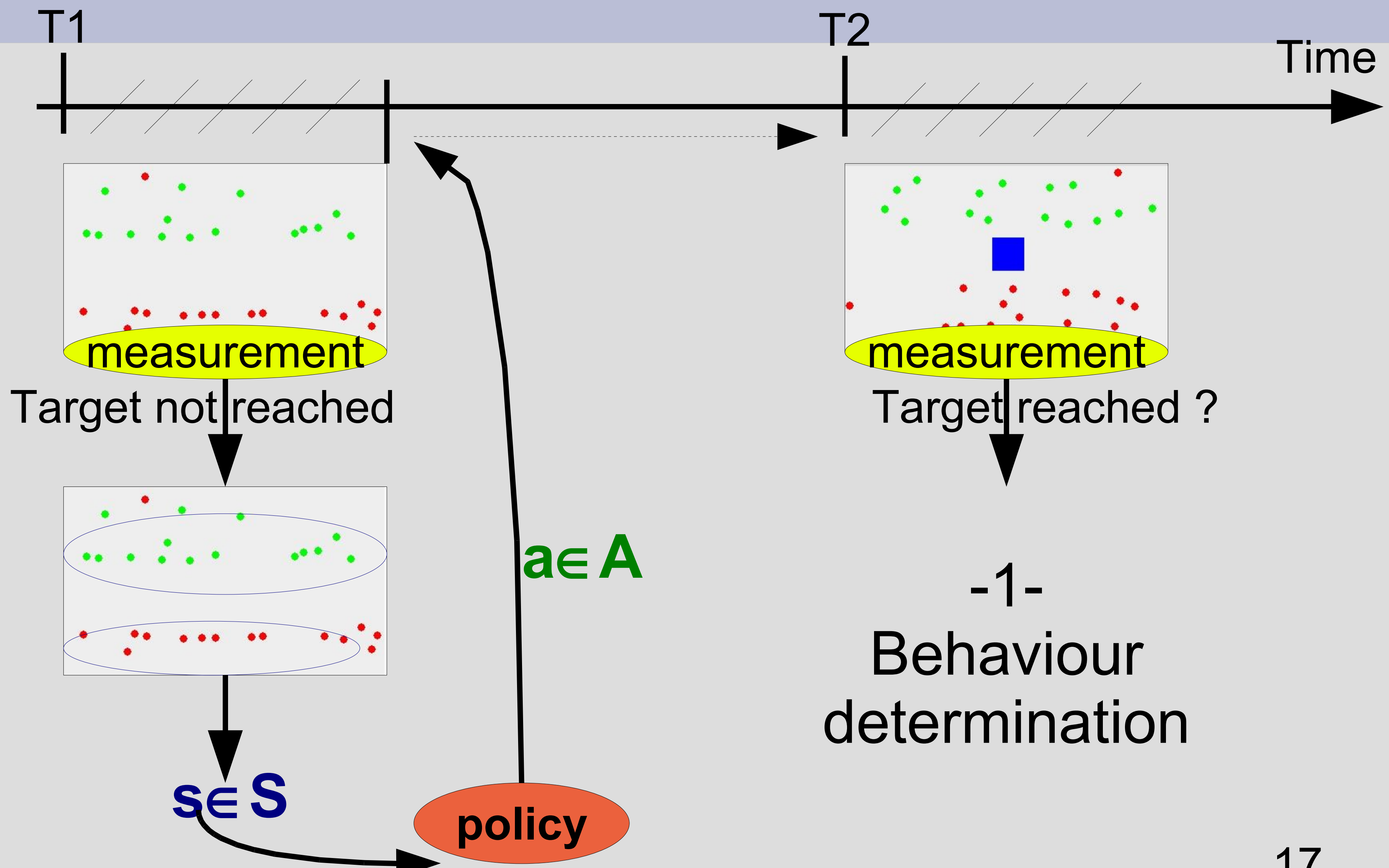
Summary


T1
Time
measurement
Target not reached
a∈ A
s∈ S
policy
-3-
Action
decision

Summary


T1
T2
Time
measurement
Target not reached
a∈ A
s∈ S
policy
-4-
Stabilisation

Summary


T1
T2
Time
measurement
Target not reached
measurement
Target reached ?
a∈A
s∈S
policy
-1-
Behaviour
determination

Case study and assessment

- Application to the toy example
  - 4 steps method
  - Applied to the pedestrians system
  - Control target : number of lines and blocks
- Assessment of the application of the method
  - Results on 2 scenarios
- Discussion
  - Assessment of the method

Application to the toy example (1)

- Global behaviour measure **measurement**
  - Number of lines and blocks
  - Clustering problem, unknown number of clusters
    Partially decentralised algorithm
- Learning of the control policy **policy**
  - Stochastic policy
    to prevent the system from staying in an attractor
  - Sarsa algorithm over 3000 simulations
    up to 50 actions in each one

Application to the toy example (2)

- States definition **S**
  - Number of lines and blocks (= global behaviour)
  - 18 different states
- Control actions **A**
  - Individual behaviours modification
    - Identical for all the agents
  - Choice between 5 values for 2 or 3 parameters
    - Coefficient of movement force
    - Coefficient of separation force
    - (Maximum speed)

Assessment

- System's controlability verification
  - Control improvement by the method ?
- Proposition compared to 2 other policies
  - Random policy
    - A random action is chosen each time a state is identified
  - Dynamical application of parameter setting
    - A *best* action a is found after evaluating each one
    - The action a is alternatively applied with a random action

Results on 2 scenarios

- Evaluation of
  - cv : rate of convergence toward the target
  - nbA : average number of actions before the target is reached

	1st scenario	*2nd scenario*
Target	1 block and 2 lines	0 block and 2 lines
Actions	25 possible actions (2 parameters)	125 possible actions (3 parameters)

Results on 2 scenarios

- Evaluation of
  - cv : rate of convergence toward the target
  - nbA : average number of actions before the target is reached

Method	*1st scenario*		*2nd scenario*	
	cv	nbA	cv	nbA
Random method	69%	15	23%	15
Parameter setting	89%	12	48%	7
Proposed method	94%	8	66%	13

Discussion

- Implementation
  - Improvement of control efficiency
  - For the studied MAS, ∃ sets **A** & **S** at a global level such as they improve the control assessment
- Method
  - Allows an effective control
  - Learning in a reasonable time / number of simulations

Conclusion and future works Proposition

- Control method
- 4 key steps
  - Global behaviour measurement
  - States description
  - Possible actions decision
  - Policy computation (reinforcement learning)

(Global behaviour measurement, States description, Possible actions decision: System dependent)

Conclusion and future works Synthesis and advantages

- Dynamical approach
  - Choice of an action in **A**
  - Depending on the state in **S**
- Automatic policy computing
- Observed global regularities can be used to improve the control efficiency
  - The controller can navigate from one state (or one global behaviour) to another

Future works

- Make the implementation more decentralised
  - In the presented implementation
    - Use of global information (global behaviour)
    - To change the behaviours of all the agents
  - Use of local information (different choice of **S**)
    - Example: an agent can be in 2 states, wether it belongs
      - to a line
      - to a block
  - Different choice of **A**
    - Examples: actions on environment or on luring agents

Questions ?