Beyond obscurantism and illegitimate curiosity

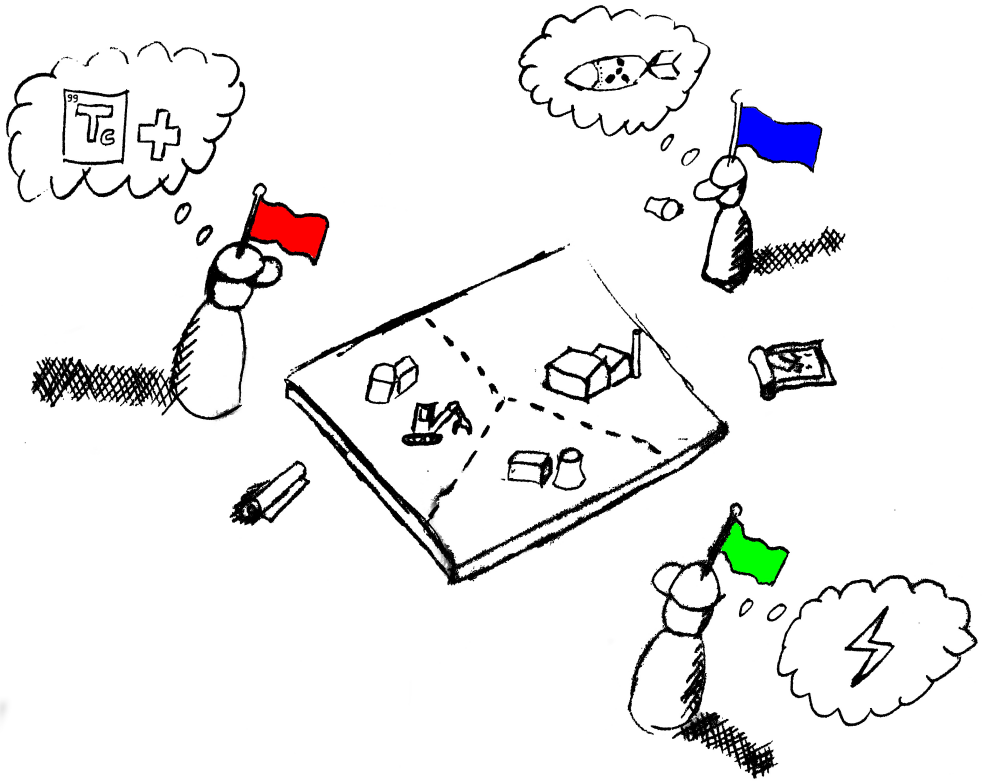
how to be transparent only with a restricted set of trusted agents

TU Delft

Nicolas Cointe and Amineh Gorbhani and Caspar Chorus

May 14th 2019

Context and aim



Cognitive agents performing *actions* in a shared *environment* to achieve their *goals*

- How to identify the most likely goals of the others?
- How to hide/show our own goals?

I do not talk about deception or proof

Introduction

Plan recognition and decision making in multiagent systems

Agent

An *entity* performing *actions* in *environment*
e.g.: human, software, robot, ...

Behavior

An *ordered* set of *actions*
e.g.: $b_{ag_i} = \{Go\ to\ ground\ floor,\ Order\ a\ capuccino,\ Drink\ it\}$

Introduction

Who is interested in goal inference ?

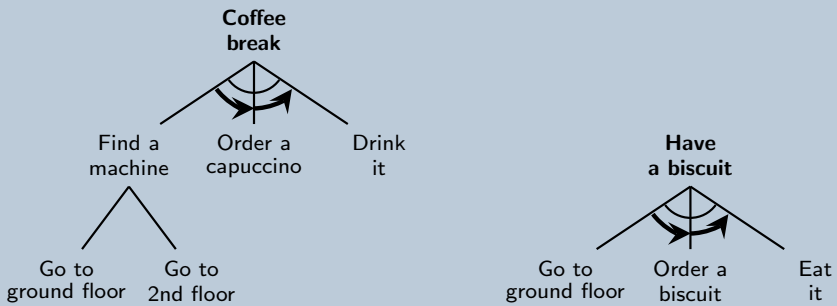
Introduction

Who is interested in goal inference ?

Plan library

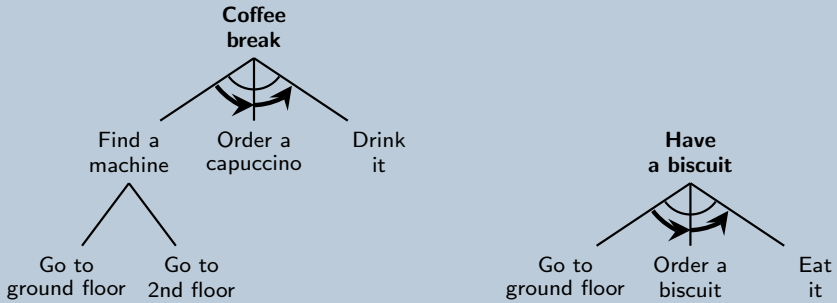
A *plan library* is a forest of *plan trees*, where roots are *intendable goals* and leaves are *actions*

e.g.:



Plan library

A *plan library* is a forest of *plan trees*, where roots are *intendable goals* and leaves are *actions*
e.g.:

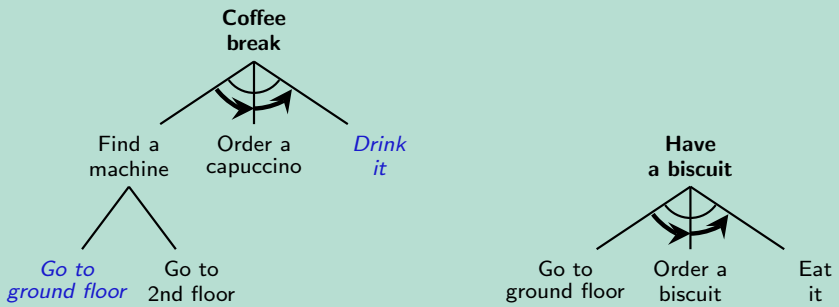


explanation

An *explanation* associates to every actions of a behavior a goal, according with the constraints.

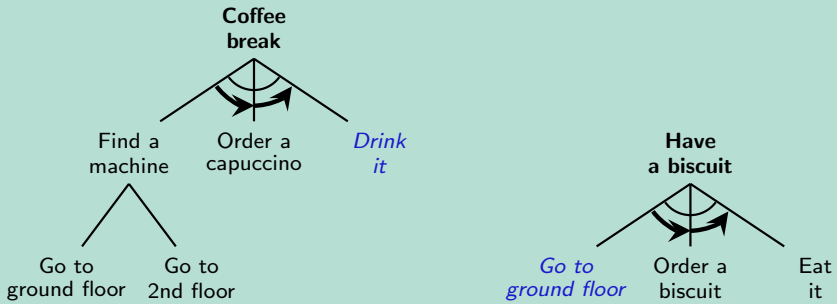
Example of explanation of a behavior

e.g.: { 9h30 go to ground floor, 9h40 drink a cappuccino }



Example of explanation of a behavior

e.g.: { 9h30 go to ground floor, 9h40 drink a cappuccino }



A behavior may have several explanations according with a plan library

Plan recognition

A *plan recognition (PR)* algorithm associates to a tuple $\langle L, b \rangle$ a *Most Probable Goal* (MPG)

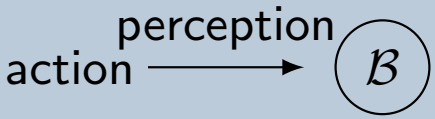
Someone is observing the elevator and I want a coffe...

- Transparency: "Where should I go to make my goal as obvious as possible ?"
- Obfuscation: "Where should I go to hide my goal as much as possible ?"

Contribution

Design of an onlooker

Analyzing behaviors (for an onlooker)

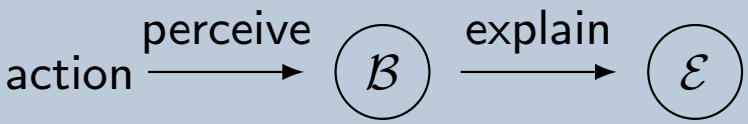


1. When an action is observed, the behavior of the acting agent is updated

Contribution

Design of an onlooker

Analyzing behaviors (for an onlooker)

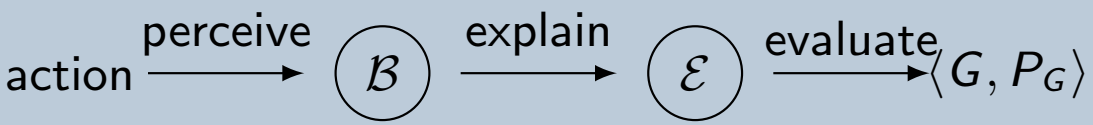


1. When an action is observed, the behavior of the acting agent is updated
2. An *explanation function* generates the set $\mathcal{E}$ of the possible explanations

Contribution

Design of an onlooker

Analyzing behaviors (for an onlooker)



1. When an action is observed, the behavior of the acting agent is updated
2. An *explanation function* generates the set $\mathcal{E}$ of the possible explanations
3. Evaluate the probability p_g of each goal $g \in G$ (proportion of g in $\mathcal{E}$)

Contribution

Design of an onlooker

Analyzing behaviors (for an onlooker)

1. When an action is observed, the behavior of the acting agent is updated
2. An *explanation function* generates the set $\mathcal{E}$ of the possible explanations
3. Evaluate the probability p_g of each goal $g \in IG$ (proportion of g in $\mathcal{E}$)
4. Pick the Most Probable Goal

Contribution

Design of an onlooker

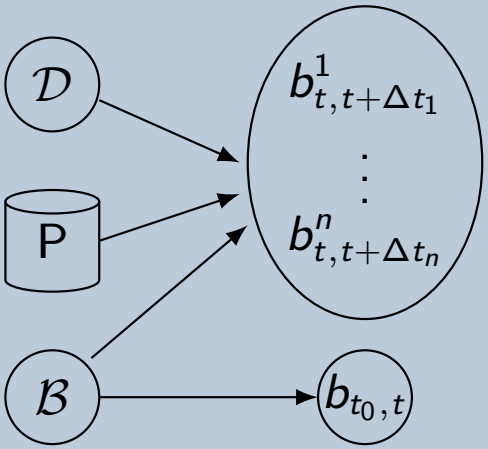
Intuition

An obfuscator/transparent agent is observing her own behavior, and analyzes the different possible choices as if they were already executed and observed (counterfactuals)

Contribution

Incorporating obfuscation and transparency in the decision

Obfuscation-based (or transparency-based) decision making



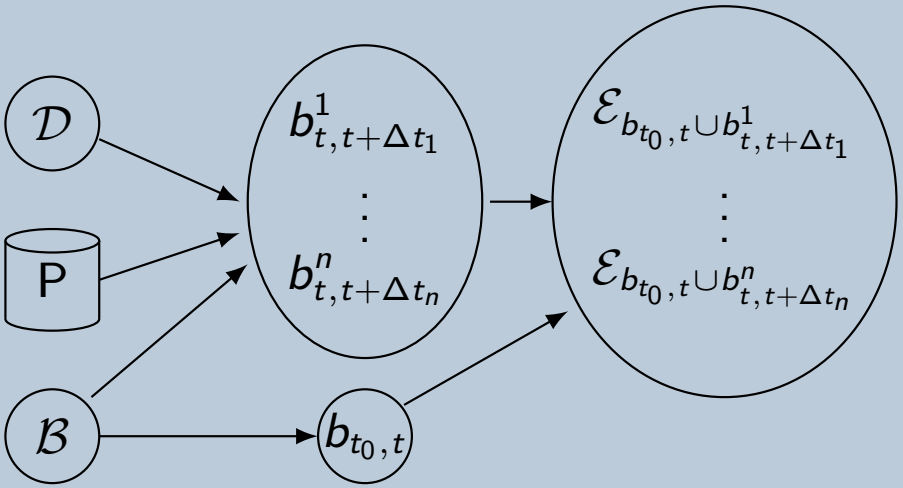
The agent knows the situation, including her past behavior (through $\mathcal{B}$), a set of desires and a set of plans

1. Infer the sequences of actions (plan execution) leading to achieve a goal

Contribution

Incorporating obfuscation and transparency in the decision

Obfuscation-based (or transparency-based) decision making

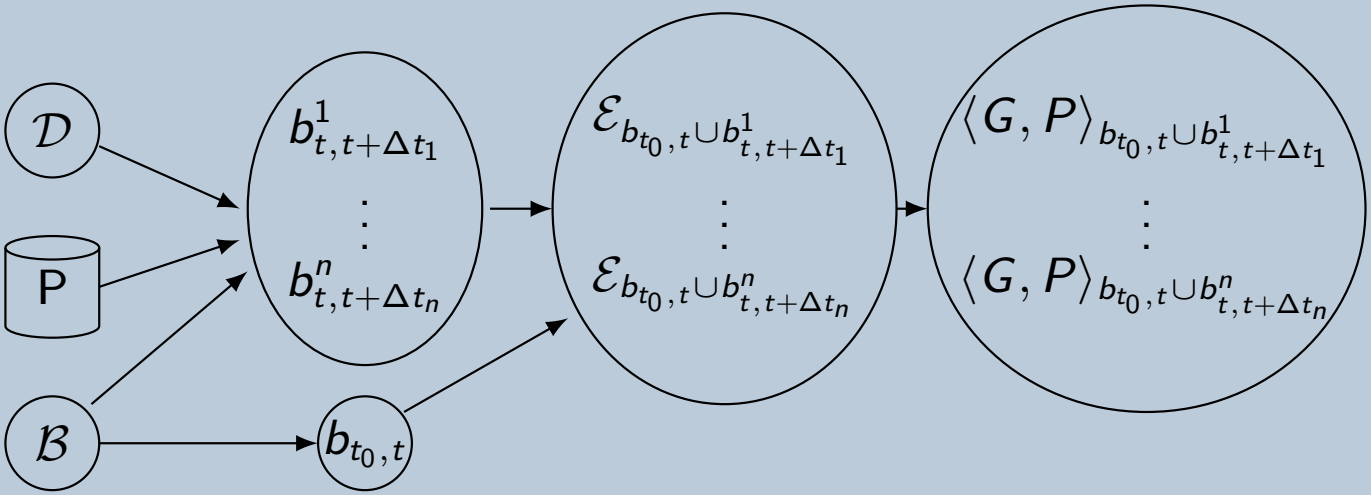


1. Infer the sequences of actions (plan execution) leading to achieve a goal
2. Evaluate the set of possible explanations for each resulting behavior

Contribution

Incorporating obfuscation and transparency in the decision

Obfuscation-based (or transparency-based) decision making

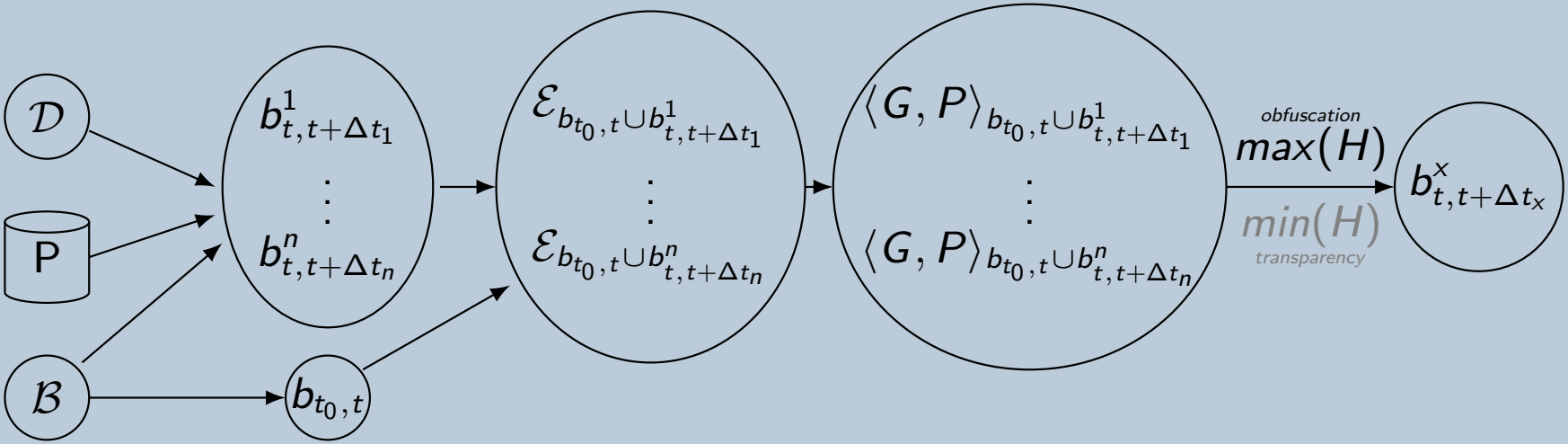


1. Infer the sequences of actions (plan execution) leading to achieve a goal
2. Evaluate the set of possible explanations for each resulting behavior
3. Compute the probability of each intendable goal for each behavior

Contribution

Incorporating obfuscation and transparency in the decision

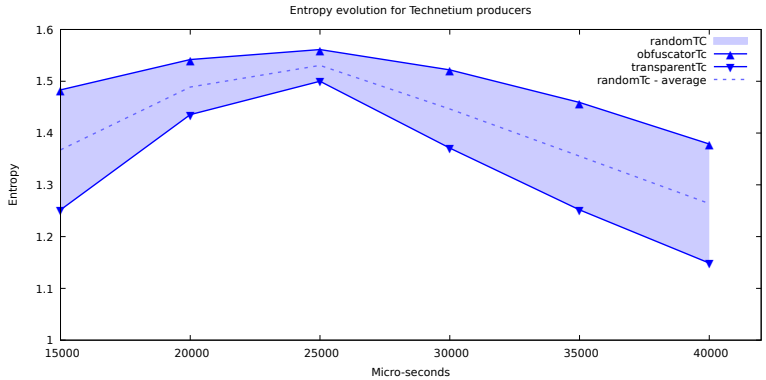
Obfuscation-based (or transparency-based) decision making

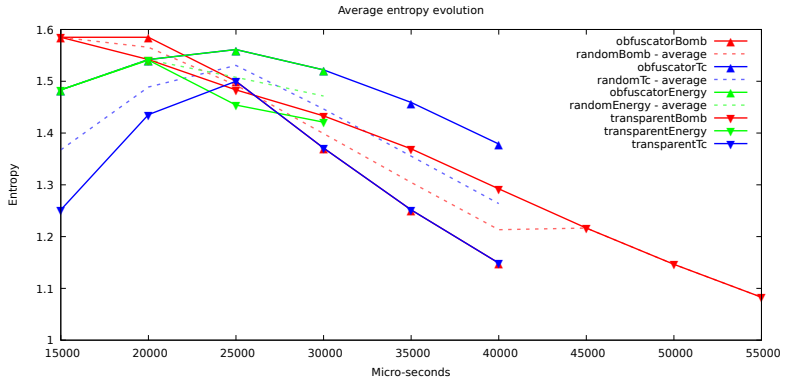


1. Infer the sequences of actions (plan execution) leading to achieve a goal
2. Evaluate the set of possible explanations for each resulting behavior
3. Compute the probability of each intendable goal for each behavior
4. The best decision maximizes (minimizes) the entropy in the probabilities

Proof of concept

Evaluation





Several experiments with realistic scenarios or randomly generated plan libraries.
Using JaCaMo.
Download run it from `http://www.nicolascointe.eu/projects/`

Collaboration and obfuscation

I am your new collaborator

Being in a coalition with obfuscators might be freaky and make trust more complicated

- Problem of *legibility* (Kirsch, 2017)
- Need of *contrastive* explanations (Winikoff, this morning)

Collaboration and obfuscation

I am your new collaborator

Being in a coalition with obfuscators might be freaky and make trust more complicated

- Problem of *legibility* (Kirsch, 2017)
- Need of *contrastive* explanations (Winikoff, this morning)

An obfuscator is still able to provide the list of subgoals from an action to the intendable goal.

Conclusion and perspectives

Transparency $\neq$ Explainability

Results

- Agents are now able to evaluate the most interesting option to hide/show their goals
- Genericity of the model
- Exhaustive exploration
- More diversity (realism ?) in simulations

Next steps

- What is the impact of partial observability ?
  - $\rightarrow$ to implement randomness/spatial limitation in the perception function ?
- Implement alternative decision making process ?
  - $\rightarrow$ Real time optimization
- Define a framework to allow obfuscation-based coalition making

Thank you for your questions and please, visit Saguenay