

Collaborative Artificial Player

Olivier Cailloux

LAMSADE, Université Paris-Dauphine

6th December, 2016

Goal

- Usually: AI VS player
- Here: AI helps player
- Illustration: chess
- Hard for a human to interpret AI moves
- Long-term goal: Develop a shared understanding of the game
- Short-term goal: Build an agent that helps human improve

How?

- Agent helps playing: suggests a move, explains why
- Test (rough idea): human applies the move
- Better idea: agent must choose a strategy among a restricted set
- Example: values of the pieces
- Mandates generalization ability
- Human sees the explanation
- Test: human tends to win the game when helped
- Link to explanations in decision theory?

Motivation

- Current: AI in its own world ([Finn Brunton](#))
- Expert \Rightarrow AI
- Similar to: oracle that understands the world
- Knows the Answer to the Ultimate Question of Life, the Universe, and Everything
- Desired: shared understanding ([David Deutsch](#))
- Expert \Rightarrow AI \Rightarrow Beginner (/ Expert)

Thank you for your attention!

Bibliography I



