Social Planning: *Achieving Goals by Altering Others' Mental States*

Chris Pearce
Ben Meadows
Pat Langley
Mike Barley

Department of Computer Science
The University of Auckland

Thanks to Miranda Emery, Alfredo Gabaldon, and Trevor Gee for their contributions to this research, which was partly funded by ONR Grant No. N00014-10-1-0487.

Research on Cognitive Systems

The cognitive systems paradigm differs from mainstream AI in that it:

- Focuses on *high-level* cognition;
- Adopts *structured representations*;
- Takes a *systems* perspective on intelligence;
- Draws inspiration from results on *human cognition*;
- Relies on *heuristic* and *satisficing* methods; and
- Encourages *exploratory research* on novel problems.

Early AI research had similar features, which makes cognitive systems closer to the field's original spirit.

See *Advances in Cognitive Systems* (http://www.cogsys.org/).

Cognitive Systems and Social Planning

Humans produce plans in many social settings with little effort; we can easily generate social plans that refer to:

- The beliefs and goals of other agents;
- Their awareness / ignorance of the true situation;
- Their reasons for carrying our social actions; and
- Even their intentions to deceive other agents.

This planning ability is an important facet of human intelligence and thus a natural target for cognitive systems research.

Definition of Social Planning

We can define the task of social planning more precisely as:

- *Given:* An initial situation, including others' mental states;
- *Given:* A set of physical or mental goals to be achieved;
- *Given:* A set of physical and *communicative* actions, along with their conditional effects on the world and others;
- *Find:* A situation that satisfies the goals and a sequence of actions that produces it.

This task is similar to the standard problem of plan generation.

But an agent must not only represent and reason about its own beliefs and goals, but about *others'* beliefs and goals.

Social Planning in Fable Settings

We can explore social planning in fable-like settings that involve interacting agents in simple environments:

> *Lion Fools the Sheep.* An aging lion is at his cave near a field. He is hungry but he can no longer chase down prey. He sees a sheep in the field but also knows it considers him dangerous. So he tells the sheep that he is ill and invites it to visit. The sheep believes the lion is harmless and comes to the cave, where the lion devours it, satisfying his hunger.

Here the lion uses communicative actions to alter the sheep's beliefs / goals, which help the lion achieve his own goals.

A Flexible Problem Solver

In previous work, we have developed FPS, a problem-solving architecture that supports different strategies for:

- Search organization (*depth first*, *breadth first*, *iterative sampling*)
- Operator selection (*means-ends analysis*, *forward search*)
- Operator application (*eager*, *delayed commitment*)
- Failure recognition (*depth limited*, effort limited, *loops*)
- Success recognition (*single*, multiple, all)

Experience with FPS's flexibility encouraged us to adapt it to support social planning.

The SFPS System

We have augmented the FPS system on a number of dimensions, including:

- Extending its representation to incorporate:
  - Others' mental states in problem states and goals;
  - Social operators that alter these mental states;
- Extending its mechanisms to:
  - Generate embedded inferences about others' beliefs;
  - Select social operators based on their main effects.

These let the resulting system – SFPS – generate social plans that involve manipulating others' mental states.

Social Operators

Our social operators involve communicative actions that alter others' mental states:

Bluff(A1, A2, Content) [A1 = actor]
  at(A1, Place), at(A2, Place)
  belief(A2, not(Content))
  belief(A1, belief(A2, not(Content))
  not(belief(A2, belief(A1, not(Content))))
  belief(A1, not(belief(A2, belief(A1, not(Content)))))
=>
  belief(A2, belief(A1, Content))
  belief(A1, belief(A2, belief(A1, Content)))

Here the acting agent (A1) tells another agent (A2) something A2 does not believe to cause A2 to believe A1 has a false belief.

Main Effects and Side Effects

Social operators have main effects and optional side effects that occur incidentally:

Bluff(A1, A2, Content) [A1 = actor]
  at(A1, Place), at(A2, Place)
  belief(A2, not(Content))
  belief(A1, belief(A2, not(Content))
  not(belief(A2, belief(A1, not(Content))))
  belief(A1, not(belief(A2, belief(A1, not(Content)))))
=>
  belief(A2, belief(A1, Content)) **[main effect]**
  belief(A1, belief(A2, belief(A1, Content))) **[side effect]**

Selection of intentions during search favors operators that use main effects to achieve goals.

Embedded Inference

Scenarios that involve interaction often require agents to reason about each others' mental states.

- One way that SFPS accomplishes this feat is through a form of *embedded inference.*
- This applies agent A's inference rules in a model of another agent B's mental state to infer B's beliefs and goals.

Because it allows reasoning about operators' indirect effects on others, this ability is vital to SFPS's social planning.

A Sample Plan

Let us return to a scenario we described earlier, *Lion Fools the Sheep.* Consider the target plan:

deceive(lion, sheep, sick(lion))
persuade(lion, sheep, at(sheep, cave))
travel(sheep, field, cave)
kill_and_eat(lion, sheep)

SFPS generates this plan by drawing on both communicative and physical operators.

The *deceive* operator creates a false belief that enables later use of *persuade* to encourage adoption of a goal.

Empirical Claims About SFPS

We make three claims about our extensions to FPS to support social planning:

- The system can create plausible plans for achieving goals in social scenarios;
- This ability relies on embedded inference to generate models of others' mental states; and
- This ability also relies on SFPS's capacity to incorporate the actions of other agents into its plans.

We designed and carried out experiments designed to test each of these claims.

Basic Results on Social Planning

We ran SFPS on eight social planning scenarios that require different levels of sophistication. Each cell gives the average of 50 runs.

Level of Sophistication	Plausible Plan	Implausible Plan	Did Not Finish
Basic Social Interaction	75	25	0
Capitalize on Misbeliefs	78	22	0
Deceive Other Agents	68	29	3
Encourage False Beliefs	86	10	4

Here plans are ‘implausible’ if they involve non-primary agents performing actions to achieve their side effects.

Results from Lesion Studies

We used lesion studies on the same scenarios to test the benefits of embedded inference and incorporating other agents' actions.

- Without embedded inference, SFPS found plausible plans on only 48 percent of its runs.
- When plans could not include actions by nonprimary agents, it found plans on only 13 percent, all on single problem.

These results are not surprising, but they provide a sanity check that the extensions are crucial to social planning.

Related Research

Our approach relies on three assumptions that have been explored in previous research:

- *Social planning relies on encoding the primary agent's beliefs and goals about others' mental states and operators for altering them.*
  - Perrault/Allen (1980), Levesque et al. (1990), Briggs/Scheutz (2013)
- *Social planning benefits from inference about problem states, which includes application of rules at different levels of embedding.*
  - Bello (2011), Fahlman (2011), Bridewell and Isaac (2011)
- *Social planning incorporates other agents' actions into plans, but only in a constrained way that avoids wishful thinking.*
  - Meehan (1977), Riedl and Young (2010)

Our work incorporates ideas from these earlier traditions, but it combines them in novel ways to support social planning.

Concluding Remarks

The task of social planning involves altering others' mental states and influencing their actions; this requires

- Representing other agents' beliefs and goals;
- Encoding social operators that alter others' mental states;
- Carrying out inference about these mental states;
- Selecting operators based on main rather than side effects.

Preliminary studies with SFPS support this analysis, but one could extend other planners in the same manner.

In future work, we should extend the framework to incorporate *abduction* and to reason about others' *emotions* and *dispositions*.

End of Presentation