

# A Game Theoretic Approach to Ad-hoc Coalitions in Human-Robot Societies

Tathagata Chakraborti Venkata Vamsikrishna Meduri Vivek Dondeti Subbarao Kambhampati

Department of Computer Science

Arizona State University

Tempe, AZ 85281, USA

{tchakra2, vmeduri, vdondeti, rao}@asu.edu

## Abstract

As robots evolve into fully autonomous agents, settings involving human-robot teams will evolve into human-robot societies, where multiple independent agents and teams, both humans and robots, coexist and work in harmony. Given such a scenario, the question we ask is - *How can two or more such agents dynamically form coalitions or teams for mutual benefit with minimal prior coordination?* In this work, we provide a game theoretic solution to address this problem. We will first look at a situation with full information, provide approximations to compute the extensive form game more efficiently, and then extend the formulation to account for scenarios when the human is not totally confident of its potential partner's intentions. Finally we will look at possible extensions of the game, that can capture different aspects of decision making with respect to ad-hoc coalition formation in human-robot societies.

Robots are increasingly becoming capable of performing daily tasks with accuracy and reliability, and are thus getting integrated into different fields of work that were until now traditionally limited to humans only. This has made the dream of human-robot cohabitation a not so distant reality. In this work we envisage such an environment where humans and robots participate autonomously (possibly with required interactions) with their own set of tasks to achieve. It has been argued (Chakraborti et al. 2016) that interactions in such situations are inherently different from those studied in traditional human-robot teams. One typical aspect of such interactions is the lack of prior coordination or shared information, due to the absence of an explicit team.

This brings us to the problem we intend to address in this paper - given a set of tasks to achieve, how can an agent proceed to select which one to achieve? In a shared environment such as the one we described, this problem cannot be simply solved by picking the goal with the highest individual utility, because the utility, and sometimes even the success of the plan (and hence the corresponding goal) of an agent are contingent on the intentions of the other agents around it. However, such interactions are not adversarial - it is just that the environment is shared among self-interested agents. Thus, an agent may choose to form an ad-hoc team with another agent in order to increase its utility, and such coalition formation should preferably be feasible with minimum prior

coordination. For example, a human with a goal to deliver two items to two different locations may team up with a delivery robot that can accomplish half of his task. Further, if the robot was itself going to be headed in one of those directions, then it is in the interest of both these agents to form this coalition. However, if the robot's plan becomes too expensive as a result, it might decide that there is not enough incentive to form this coalition. Moreover, as we highlighted before, possible interactions between agents are not just restricted to cooperative scenarios only - the plans of one agent can make the other agent's plans fail, and it may happen that it is not feasible at all for all agents to achieve their respective goals. Thus there are many possible modes of interaction between such agents, some cooperative and some destructive, that needs to be accounted for before the agents can decide on their best course of action - both in terms of which goal to choose and how to achieve it.

In this paper we model this problem of optimal goal selection as a two player game with perfect information, and propose to cut down on the prior coordination of forming such ad-hoc coalitions by looking for Nash equilibriums or socially optimal solutions (because neither agent participating in such a coalition would have incentive to deviate). We subsequently extend it to a Bayesian game to account for situations when agents are not sure of each other's intent. We will also look at properties, approximations, and interesting caveats of these games, and motivate several extensions that can capture a wide variety of ad-hoc interactions.

## 1 Related Work

There is a huge variety of work that looks at team formation from different angles. The scope of our discussion has close ties with concepts of required cooperation and capabilities of teams to solve general planning problems, introduced in (Zhang and Kambhampati 2014), and work on team formation mechanisms and properties of teams (Shoham and Tennenholtz 1992; Tambe 1997). However, in this particular work, we are more interested in the mechanism of choosing goals that can lend to possible cooperative interactions, as opposed to the mechanism of team design based on the goals themselves. Thus the work of Zhang and Kambhampati can provide interesting heuristics towards cutting down on the computation of the extensive form game we will propose, while existing work on different modes of team formation

contribute to the motivation of the Bayesian formulation of the game discussed in later sections.

From the game theoretic point of view, coalition formation has been a subject of intense study (Ray and Vohra 2014) and the human-robot interaction community can derive significant insights from it. Of particular interest are Overlapping Coalition Formation or OCF Games (Zick, Chalkiadakis, and Elkind 2012; Zick and Elkind 2014), which look at a cooperative game where the players are endowed with resources, with provisions for the players to display different modes of coalitions based on how they utilize the resources. OCF games use arbitration functions that decide the payoffs for the deviating players based on how it is affecting the non-deviating players and it helps in forming stable coalitions. This becomes increasingly relevant in shared environments such as the one we discuss here. Finally, an interesting problem that can often occur is such situations (especially with the way we have formulated the game in the human’s favor) is the problem of free-riding where agents take advantage of coalitions and try to minimize their effort (Ackerman and Brânzei 2014), which is certainly an important aspect of designing such games.

## 2 Preliminaries

### 2.1 Environment and Agent Models

**Definition 1.0** The environment is defined as a tuple  $\mathcal{E} = \langle \mathbb{F}, \mathbb{O}, \Phi, \mathcal{G}, \Lambda \rangle$ , where  $\mathbb{F}$  is a set of first order predicates that describes the environment, and  $\mathbb{O}$  is the set of objects,  $\Phi \subseteq \mathbb{O}$  is the set of agents (which may be humans or robots),  $\mathcal{G} = \{g \mid g \subseteq \mathbb{F}_0\}$ <sup>1</sup> is the set of goals that these agents are tasked with, and  $\Lambda \subseteq \mathbb{O}$  is the set of resources. Each goal has a reward  $R(g) \in \mathbb{R}^+$  associated with it.

We use PDDL (Mcdermott et al. 1998) style agent models for the rest of the discussion, but most of the analysis easily generalizes to other modes of representation. The domain model  $D_\phi$  of an agent  $\phi \in \Phi$  is defined as  $D_\phi = \langle \mathbb{F}_0, A_\phi \rangle$ , where  $A_\phi$  is a set of operators available to the agent. The action models  $a \in A_\phi$  are represented as  $a = \langle \mathbb{C}_a, \mathbb{P}_a, \mathbb{E}_a \rangle$  where  $\mathbb{C}_a$  is the cost of the action,  $\mathbb{P}_a \subseteq \mathbb{F}_0$  is the list of pre-conditions that must hold for the action  $a$  to be applicable in a particular state  $S \subseteq \mathbb{F}_0$  of the environment; and  $\mathbb{E}_a = \langle \text{eff}^+(a), \text{eff}^-(a) \rangle$ ,  $\text{eff}^\pm(a) \subseteq \mathbb{F}_0$  is a tuple that contains the add and delete effects of applying the action to a state. The transition function  $\delta(\cdot)$  determines the next state after the application of action  $a$  in state  $S$  as  $\delta(a, S) \models \perp$  if  $\mathbb{P}_a \not\subseteq S$ ;  $\models (S \setminus \text{eff}^-(a)) \cup \text{eff}^+(a)$  otherwise.

A planning problem for the agent  $\phi$  is given by the tuple  $\Pi_\phi = \langle \mathbb{F}_0, D_\phi, \mathbb{I}_\phi, \mathbb{G}_\phi \rangle$ , where  $\mathbb{I}_\phi, \mathbb{G}_\phi \subseteq \mathbb{F}_0$  are the initial and goal states respectively. The solution to the planning problem is an ordered sequence of actions or *plan* given by  $\pi_\phi = \langle a_1, a_2, \dots, a_{|\pi_\phi|} \rangle$ ,  $a_i \in A_\phi$  such that  $\delta(\pi_\phi, \mathbb{I}_\phi) \models \mathbb{G}_\phi$ , where the cumulative transition function is given by  $\delta(\pi, s) = \delta(\langle a_2, a_3, \dots, a_{|\pi|} \rangle, \delta(a_1, s))$ . The cost of the plan is given by  $C(\pi_\phi) = \sum_{a \in \pi_\phi} \mathbb{C}_a$  and the optimal plan  $\pi_\phi^*$  is such that  $C(\pi_\phi^*) \leq C(\pi_\phi) \forall \pi_\phi$  with  $\delta(\pi_\phi, \mathbb{I}_\phi) \models \mathbb{G}_\phi$ .

<sup>1</sup> $S_0$  is any  $S \subseteq \mathbb{F}$  instantiated / grounded with objects from  $\mathbb{O}$ .

### 2.2 Representation of Human-Robot Coalitions

We will represent coalitions of such agents by means of a super-agent transformation (Chakraborti et al. 2015a) on a set of agents that combines the capabilities of one or more agents to perform complex tasks that a single agent might not be capable of doing. Note that this does not preclude joint actions among agents, because some actions that need that need more than one agent (as required in the preconditions) will only be doable in the composite domain.

**Definition 1.1** A super-agent is a tuple  $\Theta = \langle \theta, D_\theta \rangle$  where  $\theta \subseteq \Phi$  is a set of agents in the environment  $\mathcal{E}$ , and  $D_\theta$  is the transformation from the individual domain models to a composite domain model given by  $D_\theta = \langle \mathbb{F}_0, \bigcup_{\phi \in \theta} A_\phi \rangle$ .

**Definition 1.2** The planning problem of a super-agent  $\Theta$  is given by  $\Pi_\Theta = \langle \mathbb{F}_0, D_\theta, \mathbb{I}_\theta, \mathbb{G}_\theta \rangle$  where the composite initial and goal states are given by  $\mathbb{I}_\theta = \bigcup_{\phi \in \theta} \mathbb{I}_\phi$  and  $\mathbb{G}_\theta = \bigcup_{\phi \in \theta} \mathbb{G}_\phi$  respectively. The solution to the planning problem is a composite plan  $\pi_\theta = \langle \mu_1, \mu_2, \dots, \mu_{|\pi_\theta|} \rangle$  where  $\mu_i = \{a_1, \dots, a_{|\theta|}\}$ ,  $\mu(\phi) = a \in A_\phi \forall \mu \in \pi_\theta$  such that  $\delta'(\mathbb{I}_\theta, \pi_\theta) \models \mathbb{G}_\theta$ , where the modified transition function  $\delta'(\mu, s) = (s \setminus \bigcup_{a \in \mu} \text{eff}^-(a)) \cup \bigcup_{a \in \mu} \text{eff}^+(a)$ . The cost of a composite plan is  $C(\pi_\theta) = \sum_{\mu \in \pi_\theta} \sum_{a \in \mu} \mathbb{C}_a$  and  $\pi_\theta^*$  is optimal if  $C(\pi_\theta^*) \leq C(\pi_\theta) \forall \pi_\theta$  with  $\delta'(\mathbb{I}_\theta, \pi_\theta) \models \mathbb{G}_\theta$ . The composite plan can be viewed as a union of plans contributed by each agent  $\phi \in \theta$  whose component can be written as  $\pi_\theta(\phi) = \langle a_1, a_2, \dots, a_n \rangle$ ,  $a_i = \mu_i(\phi) \forall \mu_i \in \pi_\theta$ .

### 2.3 The Use Case

Throughout the rest of the discussion we will use the setting from Talamadupula et al. which involves a human commander `CommX` and a robot in a typical Urban Search and Rescue (USAR) scenario, as illustrated in Figure 1. The environment consists of interconnected rooms and hallways, which the agents can navigate and search. The commander can perform triage in certain locations, for which he needs the medkit. The robot can also fetch medkits if requested by other agents (not shown) in the environment. A sample domain is available at <http://bit.ly/1Fko7MA>. The shared resources here are the two medkits - i.e. some of the plans the agents can execute will lock the use of and/or change the position of these medkits, so as to make the other agent’s plans, contingent on that particular resource, invalid.

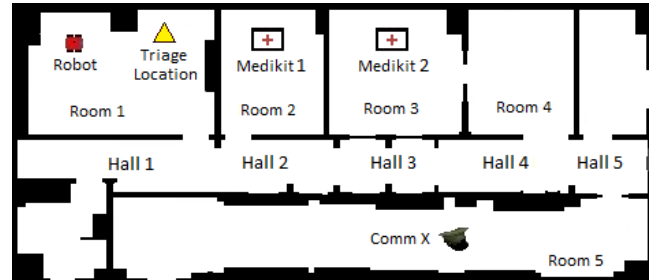


Figure 1: Use case - Urban Search And Rescue (USAR).

### 3 Ad-hoc Human-Robot Coalitions

In this section we will look at how two agents (the human and the robot) in our scenario, can coordinate dynamically by forming impromptu teams in order to achieve either individually rational or socially optimal behaviors.

#### 3.1 Motivation

Consider the scenario shown in Figure 1. Suppose one of CommX’s goal is to perform triage in room1, while one of the Robot’s goals is to deliver a medkit to room1. Clearly, if both the agents choose to do their optimal plans and plan to use medkit1 in room2, the Robot’s plan fails (assuming the CommX gets there first). The robot then has two choices - (1) it can choose to achieve some other goal, i.e. maximize it’s own rewards, (2) it can choose to deliver the other medkit2 from room3, i.e. maximize social good.

Indeed there are many possible ways that these agents can interact. For example, the utility of choosing any goal may be defined by the optimal cost of achieving that goal individually, or as a team. This in turn affects the choice whether to form such teams or not. In the discussion that follows, we model this goal selection (and team formation) problem as a strategic game with perfect information.

#### 3.2 Formulation of the Game

We refer to our static two-player strategic game Goal Allocation with Perfect Information as  $GAPI = \langle \Phi, \{A^\phi\}, \{U^\phi\} \rangle$ . The game attempts to determine, given complete information about the domain model and goals of the other agent, which goal to achieve and whether forming a coalition is beneficial. The game is defined as follows -

- **Players** - The game has two players  $\Phi = \{H, R\}$  the human  $H$  and the robot  $R$  respectively.
- **Actions** - The actions of the agents in the strategic game are the *goals* that they can select to achieve. Thus, for each agent  $\phi \in \Phi$  we define a set of goals  $\mathcal{G}^\phi = \{G_\phi^1, G_\phi^2, \dots, G_\phi^{|\mathcal{G}^\phi|}\} \subseteq \mathcal{G}$ , and the action set  $A^\phi$  of the agent  $\phi$  is the mapping that assigns one of these goals as its planning goal, i.e.  $A^\phi : \mathcal{G}^\phi \mapsto \mathbb{G}$ . Note that this is distinct from the action models defined in PDDL for each of the individual agents (which helps the agent figure out *how* this goal  $\mathbb{G}$  is achieved, and the resultant utility).
- **Utilities** - Finally, as discussed previously, the utility of an action depends on (apart from the utility of the goal itself) the way the agent chooses to achieve it, and is contingent also on the plans of the other agent (due to, for example, resource conflicts), and is given by -

$$\begin{aligned} U_H(A_i^H, A_j^R) &= R(G_H^i \cup G_R^j) - \min\{C(\pi_H^*), C(\pi_\Theta^*(H))\} \\ U_R(A_i^H, A_j^R) &= R(G_H^i \cup G_R^j) - C(\pi_\Theta^*(R)) \text{ if } C(\pi_H^*) > C(\pi_\Theta^*(H)) \\ &= R(G_H^i \cup G_R^j) - \max\{C(\pi_R^*), C(\pi_\Theta^*(R))\}, \text{ otherwise.} \end{aligned}$$

where,  $\pi_H^*$  is the optimal plan or solution of the planning problem defined by  $\Pi_H = \langle \mathbb{F}_\Theta, D_H, \mathbb{I}_H, \mathbb{G}_H^i \rangle$ ,  $\pi_R^*$  is the optimal solution of  $\Pi_R = \langle \mathbb{F}_\Theta, D_R, \mathbb{I}_R, \mathbb{G}_R^j \rangle$ , and  $\pi_\Theta^*$  is the optimal solution of  $\Pi_\Theta = \langle \mathbb{F}_\Theta, D_\theta, \mathbb{I}_\theta, \mathbb{G}_\theta \rangle$ , where

$\Theta = \langle \theta, D_\theta \rangle$  is the super-agent representing the coalition formed by  $\theta = \{H, R\}$  with  $\mathbb{I}_\theta = \mathbb{I}_H \cup \mathbb{I}_R$  and  $\mathbb{G}_\theta = G_H^i \cup G_R^j$ . Here, the first term in the expression for utility denotes the utility of the goal itself as defined in the environment in Section 2.1, while the second term captures the resultant best case utility of plans due to agent interactions. More on this below.

**Human-centric robots.** At this point we make an assumption about the role of the robots in our human-robot society - we assume that the robots exist only in the capacity of autonomous assistance, i.e. in coalitions that may be formed with humans and robots, the robot’s role is to improve the quality of life of the humans (by possibly, in our case, reducing the costs of plans) and *not* vice versa.

Thus, in the expression of utility, the human uses a minimizing term - with no interactions  $C(\pi_H^*) = C(\pi_\Theta^*(H))$ , otherwise  $C(\pi_H^*) > C(\pi_\Theta^*(H))$ . Similarly, in case of the robot, with no interactions  $C(\pi_R^*) \geq C(\pi_\Theta^*(R))$  and  $C(\pi_H^*) < \Rightarrow C(\pi_\Theta^*(H))$  otherwise, since the interactions may or may not be always cooperative for the robot. Note that this formulation also takes care of the cases when the robot goal becomes unachievable due to negative interactions with the human (this is why we have the maximizing term; the difference is triggered due to negative interactions with the human plan in absence of coalitions). Also note that the goal utility is using a combined goal due to the particular action profile, this captures cases when goals have interactions, i.e. a conjunction of goals may have higher (or lower) utility than the sum of its components.

This can be easily ensured while generating plans for a given coalition, by either discounting the costs of actions of the robot with respect to those of the humans by a suitable factor, or more preferably, by just penalizing the total cost of the human component in the composite plan more. The assumption of course does not change the formulation in any way, it is just more aligned with the notion of the social robots being envisioned currently. Of course, in this sense the utilities of both the humans and robots will now become identical, with a minimizing cost term.

Now that we have defined the game, the question is how do we choose actions for each agent? Remember that we want to find solutions that will preclude the need to coordinate. We can take two approaches here - we can make agents individually rational (in which case both the human and the robot looks for a Nash equilibrium, so neither has a reason to defect; or we can make the agents look for a socially optimal solution (so that sum of utilities is maximized).

#### 3.3 Solving for Nash Equilibriums

As usual, the Nash equilibriums in  $GAPI$  are given by action profiles  $\langle A_i^H, A_j^R \rangle$  such that  $U_H(A_i^H, A_j^R) \geq U_H(A_k^H : \forall k \neq i, A_j^R)$  and  $U_R(A_i^H, A_j^R) \geq U_R(A_i^H, A_k^R : \forall k \neq j)$ . It is easy to prove that there is no guaranteed Nash equilibrium in  $GAPI$ . We will instead motivate a slightly different game  $GAPI\text{-Bounded}$  where the robot only agrees to deviate from its optimal plan up to a certain degree, i.e. there is a bound on the amount of assistance the robot chooses to provide.

**Definition 1.3.** The differential help  $\delta(g, G_R^i)$  provided by the robot  $R$  with goal  $G_R^i \in \mathcal{G}^R$ , when the human  $H$  picks goal  $g \in \mathcal{G}^H$ , measures the decrease in utility of the robot upon forming a coalition with the human, and is given by  $\delta(g, G_R^i) = |C(\pi_\Theta^*(R)) - C(\pi_R^*)|$ , where  $\pi_R^*$  is the optimal solution of  $\Pi_R = \langle \mathbb{F}_\Theta, D_R, \mathbb{I}_R, \mathbb{G}_R^i \rangle$ , and  $\pi_\Theta^*$  is the optimal solution of  $\Pi_\Theta = \langle \mathbb{F}_\Theta, D_\theta, \mathbb{I}_H \cup \mathbb{I}_R, g \cup G_R^i \rangle$ , where  $\Theta = \langle \theta = \{H, R\}, D_\theta \rangle$ .

Thus in GAPI-Bounded the utility function is modified from the one in GAPI as follows -

$$U_H(A_i^H, A_i^R) = R(G_H^i) - C(\pi_H^*)$$

$$U_R(A_i^H, A_j^R) = R(G_R^j) - C(\pi_R^*)$$

if  $\exists G_R^k : k \neq j \in \mathcal{G}^R$  s.t.  $\delta(G_H^i, G_R^j) > \{R(G_R^j) - C(\pi_R^*)\} - \{R(G_R^k) - C(\pi_R^{**})\}$ , where  $\pi_R^*$ ,  $\pi_R^{**}$  and  $\pi_H^*$  are the optimal plans or solutions to the planning problems  $\Pi_R^i = \langle \mathbb{F}_\Theta, D_R, \mathbb{I}_R, \mathbb{G}_R^i \rangle$ ,  $\Pi_R^k = \langle \mathbb{F}_\Theta, D_R, \mathbb{I}_R, \mathbb{G}_R^k \rangle$  and  $\Pi_H = \langle \mathbb{F}_\Theta, D_H, \mathbb{I}_H, \mathbb{G}_H^i \rangle$  respectively; and otherwise -

$$U_H(A_i^H, A_i^R) = R(G_H^i) - C(\pi_\Theta^*(H))$$

$$U_R(A_i^H, A_j^R) = R(G_R^j) - C(\pi_\Theta^*(R))$$

where  $\pi_\Theta^*$  is the optimal solution of  $\Pi_\Theta = \langle \mathbb{F}_\Theta, D_\theta, \mathbb{I}_H \cup \mathbb{I}_R, g \cup G_R^j \rangle$ , where  $\Theta = \langle \theta = \{H, R\}, D_\theta \rangle$ .

This basically means that if the penalty that the robot incurs by choosing to assist the human is so great that it could rather do something else instead (i.e. choose another goal), then it switches back to using its individual optimal plan, i.e. no coalition is formed. *If the individual optimal plans are always feasible* (otherwise these do not participate in the Nash equilibriums below), this leads to the following result.

**Claim.**  $\langle A_{i^*}^H, A_{j^*}^R \rangle$  must be a Nash equilibrium of GAPI-Bounded when  $j^* = \arg \max_{G_R^j \in \mathcal{G}^R} R(G_R^j) - C(\pi_R)$  and  $i = \arg \max_i U_H(G_H^i, G_R^{j^*})$ .

**Proof Sketch.** Let us define the utility function of the robot  $R$  for achieving a goal  $g \in \mathcal{G}^R$  by itself as  $\tau(g) = R(g) - C(\pi_R^*)$ , where  $\pi_R^*$  is the optimal solution to the planning problem  $\Pi_R = \langle \mathbb{F}, \mathbb{O}, D_R, \mathbb{I}_R, g \rangle$ . Further, given the goal set  $\mathcal{G}^R$  of the robot, we set  $G_R^{j^*} = \arg \max_{g \in \mathcal{G}^R} \tau(g)$ , i.e.  $G_R^{j^*}$  corresponds to the highest utility goal that the robot can achieve by itself. Now consider any two goals  $G_R^j, G_R^{j^*} \in \mathcal{G}^R$ ,  $G_R^j \neq G_R^{j^*}$ . We argue that  $\forall G_H^i \in \mathcal{G}^H$ ,  $U_R(A_i^H, A_{j^*}^R) \geq U_R(A_i^H, A_j^R)$ . This is because  $\tau(G_R^{j^*}) \geq \tau(G_R^j)$  and by problem definition  $\forall i, k | U_R(A_i^H, A_{j^*}^R) - U_R(A_i^H, A_j^R) | \leq \tau(G_R^{j^*}) - \tau(G_R^j)$ . Thus, in general, the goal ordering induced by the function  $\tau$  is preserved by the utility function  $U_R$ , and consequently  $A_{j^*}^R$  is a dominant strategy of the robot. It follows that  $A_{i^*}^H$  such that  $i^* = \arg \max_i U_H(G_H^i, G_R^{j^*})$  is the corresponding best response for the human. Hence  $\langle A_{i^*}^H, A_{j^*}^R \rangle$  must be a Nash equilibrium. Hence proved.  $\square$

Further, it may be noted here that there may be many such Nash equilibriums in GAPI-Bounded and these are also the only ones, i.e. all Nash equilibriums in GAPI-bounded must satisfy the conditions in the above claim.

### 3.4 Solving for Social Good

Similarly, the socially optimal goal selection strategies are given by the action profiles  $\langle A_{i^*}^H, A_{j^*}^R \rangle$  where  $\{i^*, j^*\} = \arg \max_{i,j} U_H(A_i^H, A_j^R) + U_R(A_i^H, A_j^R)$ . The socially optimal action profiles may not necessarily correspond to any Nash equilibriums of either GAPI or GAPI-Bounded.

**Individual Irrationality and  $\epsilon$ -Equilibrium.** Given the way the game is defined, it is easy to see that the socially good outcome may not be individually rational for either the human or the robot, since the robot always has the incentive to defect to choosing  $G_R^*$  and the human will then choose the corresponding highest utility goal for himself. This leaves room for designing autonomy that can settle for action profiles  $\langle A_{i^*}^H, A_{j^*}^R \rangle$  referred to as  $\epsilon$ -equilibriums, for the purpose of social good, i.e.  $|U_H(A_{i^*}^H, A_{j^*}^R) - U_H(A_i^H, A_j^R)| \leq \epsilon$  and  $|U_R(A_{i^*}^H, A_{j^*}^R) - U_R(A_i^H, A_j^R)| \leq \epsilon$ . Note that this deviation is distinct from the concept of bounded differential assistance we introduced in Section 3.3.

**Price of Anarchy.** The price of deviating from individual rationality is referred to as the Price of Anarchy and is measured by  $\mathcal{POS} = \frac{U_H(A_{i^*}^H, A_{j^*}^R) + U_R(A_{i^*}^H, A_{j^*}^R)}{U_H(A_i^H, A_j^R) + U_R(A_i^H, A_j^R)}$ .

### 3.5 Caveats

**No or Multiple Nash Equilibriums.** One of the obvious problems with this approach is that it does not guarantee a unique Nash equilibrium, if it exists at all. This has serious implications on the problem we set out to solve in the first place - *which goals do the agents choose to plan for, and how?* Note, however, that this is not really a feature of the formulation itself but of the domain or the environment, i.e. the action models of the agents and the utilities in the goals will determine whether there is a single best coalition that may be formed given a particular situation. Thus, there seems to be no principled way of solving this problem in a detached manner, without any form of communication between the agents. But our approach still provides a way to deliberate over the possible options, and communicate to resolve ambiguities only with respect to the Nash equilibriums, rather than the whole set of goals, or even just those in each agent's dominant strategy, which can still provide significant reduction in the communication overhead.

**Infeasibility of the Extensive Form Game.** Note here that the utilities of the actions are calculated from the cost of plans to achieve the corresponding goals, which involves solving two planning problems per action. This means that, in order to get the extensive form of GAPI, we need to solve  $\mathcal{O}(|\mathcal{G}^H| \times |\mathcal{G}^R|)$  planning problems in total (note that solving for  $\pi_\Theta$  gives utilities for both agents  $H$  and  $R$ ), which may be infeasible for large domains. So we need a way to speed up our computation (either by computing an approximation



and/or finding ways to calculate multiple utility values at once), while simultaneously preserving guarantees from our original game in our approximate version.

Fortunately, we have good news. Note that all we require are *costs* of the plans, not the plans themselves. So a promising approach towards cutting down on the computational complexity is by using heuristic values for the initial state of a particular planning problem as a proxy towards the true plan cost. Note that the better the heuristic is, the better our approximation is. So the immediate question is - What guarantees can we provide on the values of the utilities when we use heuristic approximation? Are the Nash equilibriums in the original game still preserved? This brings us to the notion of “*well-behaved heuristics*” as follows -

**Definition 1.4** A *well-behaved heuristic*  $h : S \times S \mapsto \mathbb{R}^+$ ,  $S \subseteq \mathbb{F}_0$  is such that  $h(\mathbb{I}, \mathbb{G}^1) \leq h(\mathbb{I}, \mathbb{G}^2)$  whenever  $C(\pi_1^*) \leq C(\pi_2^*)$ , where  $\pi_1^*$  and  $\pi_2^*$  are the optimal solutions to the planning problems  $\Pi_1 = \langle \mathbb{F}_0, D, \mathbb{I}, \mathbb{G}^1 \rangle$  and  $\Pi_2 = \langle \mathbb{F}_0, D, \mathbb{I}, \mathbb{G}^2 \rangle$  respectively.

We define  $\overline{\text{GAPI}}$  as a game identical to  $\text{GAPI}$  but with a modified utility function as follows -

$$\begin{aligned} U_H(A_i^H, A_j^R) &= R(G_H^i) - \min\{h(G_H^i, \mathbb{I}_H), h(G_H^i, \mathbb{I}_H \cup \mathbb{I}_R)\} \\ U_R(A_i^H, A_j^R) &= R(G_H^i) - h(G_R^j, \mathbb{I}_H \cup \mathbb{I}_R) \text{ if } h(G_H^i, \mathbb{I}_H) > h(G_H^i, \mathbb{I}_H \cup \mathbb{I}_R) \\ &= R(G_H^i) - \max\{h(G_R^j, \mathbb{I}_R), h(G_R^j, \mathbb{I}_H \cup \mathbb{I}_R)\}, \text{ otherwise.} \end{aligned}$$

Note that in order to get a heuristic estimate of an agent’s contribution to the composite plan, we compute the heuristic with respect to achieving the individual agent goal using the composite domain of the super-agent, which of course gives a lower bound on the real cost of the composite plan used to achieve that agent’s goal only.

**Claim.** NEs in  $\text{GAPI}$  are preserved in  $\overline{\text{GAPI}}$ .

**Proof Sketch.** This is easy to see because orderings among costs are preserved by a well-behaved heuristic, and hence ordering among utilities, which is known to keep the Nash equilibriums unchanged. Note that the reverse does not hold, i.e.  $\overline{\text{GAPI}}$  may have extra Nash equilibriums due to the equality in the definition of well-behaved heuristics.  $\square$

**Definition 1.5** We define a *goal-ordering on the goal set*  $\mathcal{G}^\phi$  of agent  $\phi$  as a function  $f : [1, |\mathcal{G}^\phi|] \mapsto [1, |\mathcal{G}^\phi|]$  such that  $G_\phi^{f(1)} \subseteq G_\phi^{f(2)} \subseteq \dots \subseteq G_\phi^{f(|\mathcal{G}^\phi|)}$ . This means that the goals of an agent are such that they are all different subgoals of a single conjunctive goal.

We will refer to the game with agents with such ordered goal sets as  $\overline{\overline{\text{GAPI}}}$  (identical to  $\text{GAPI}$  otherwise).

**Claim.** NEs in  $\text{GAPI}$  are preserved in  $\overline{\overline{\text{GAPI}}}$ .

**Proof Sketch.** Since  $\mathcal{G}^\phi$  is goal-ordered,  $C(\pi_{f(1)}^*) \leq C(\pi_{f(2)}^*) \leq \dots \leq C(\pi_{f(|\mathcal{G}^\phi|)}^*)$ , where, as usual,  $\pi_i^*$  is the optimal solution to the planning problem  $\Pi_i = \langle \mathbb{F}_0, D, \mathbb{I}, G_\phi^i \rangle$ . Let us consider a *non-trivial admissible heuristic*  $h$  and define a heuristic  $\hat{h}$  such that  $\hat{h}(\mathbb{I}, G_\phi^i) =$

$\max\{h(\mathbb{I}, G_\phi^i), \hat{h}(\mathbb{I}, G_\phi^{i-1})\}$ ;  $\hat{h}(\mathbb{I}, G_\phi^1) = h(\mathbb{I}, G_\phi^1)$ . Then  $\hat{h}$  is well-behaved. Hence proved.  $\square$

These properties of  $\text{GAPI}$ -Bounded,  $\overline{\text{GAPI}}$  and  $\overline{\overline{\text{GAPI}}}$  enables computation of approximations, and partial profiles, to the extensive form of  $\text{GAPI}$ , while maintaining the nature of interactions, thus making the formulation more tractable.

## 4 Bayesian Modeling of Teaming Intent

### 4.1 Motivation

In the previous sections we considered both individual and team plans, and as teams we considered optimal plans for a coalition. In reality there are many ways that a particular coalition can achieve a particular goal, and correspondingly there are different modes of interaction between the team-mates. We discuss four such possibilities briefly here -

- **Individual Optimality** - In this type of planning, each agent computes the individual optimal plan to achieve their goals. Note that this plan may not be actually valid in the environment during execution time, due to factors such as resource conflicts due to plans of the other agents.
- **Joint Optimality** - Here we compute the joint optimal for a coalition; and this optimal plan is computed in favor of the human as discussed previously in Section 3.2.
- **Planning with Resource Conflicts** - In (Chakraborti et al. 2015b) we explored a technique for the robot to produce plans so as to ensure the success of the human plans only, and explored different modes of such behavior of the robot in terms of compromise, opportunism and negotiation. Thus utilities for the human plans computed this way is, at times, same as the joint optimal, but in general is greater than or equal to the individual optimal and less than or equal to the joint optimal.
- **Planning for Serendipity** - In (Chakraborti et al. 2015a) we looked at a special case of multi-agent coordination, where the robot computes opportunities for assisting the human in the event the human is not planning to exploit the robot’s help. Here, as in the previous case, utilities for the human plans computed this way is again greater than or equal to the individual optimal and less than or equal to the joint optimal plans.

Going back to our use case in Figure 1, suppose the robot has a goal to deliver a medkit to `room1`, and `CommX` has a goal to conduct triage in `room1`, for which he also requires a medkit (and his optimal plan involves picking up `medkit1` in `room2`). For individual optimal plans both the robot and the human will go for `medkit2` (thus, in this situation, individual optimal plans are actually not feasible). For the joint optimal, the coalition can team up to both use the same medkit thus achieving mutual benefit. In case the robot is only planning to avoid conflicts, it can settle for using `medkit3` which is further away, or the robot can also intervene serendipitously by handing over `medkit2` in the hallway thus achieving higher utility through cooperation without directly coordinating.

For our problem, this has the implication that we can no longer be sure of the plan (and consequently the utility) even

when a particular goal has been chosen. Rather what we have is a possible set of utilities for each goal. However we can do better than to just take the maximum (or minimum as the case may be) of these utilities as we did previously, because we now know how such behaviors are being generated and so we can leverage additional information from an agent’s beliefs about the other agent to come up with optimal response strategies. This readily lends the problem to a formulation in terms of Bayesian strategic games, which we will discuss in the next section.

## 4.2 Formulation of the Game

We define our two-person static Bayesian game  $\text{GAPI-Bayesian} = \langle \Phi, \mathcal{B}, A^H, \{A^{R,\mathcal{B}}\}, U^H, \{U^{R,\mathcal{B}}\} \rangle$  with belief  $\mathcal{B}$  over the type of robot as follows -

- **Players** - We still have two players - the human  $H$  and the robot  $R$ , as in the previous games.
- **Actions** - The actions of the players are similarly identical to  $\text{GAPI}$ , i.e. the action set of agent  $\phi \in \{H, R\}$  is the mapping  $A^\phi : \mathcal{G}^\phi \mapsto \mathbb{G}$ .
- **Beliefs** - The human has a set of beliefs on the robot  $\mathcal{B} = \{B_1, B_2, \dots, B_{|\mathcal{B}|}\}$  characterized by the distribution  $B \sim P$ , i.e. the robot can be of any of the types in  $\mathcal{B}$  with probability  $P(B)$ . The type of the robot is essentially the *algorithm* it uses to compute the optimal plan given the initial state and the selected goal, and thus affects the cost of achieving the goal, and hence the utility function.

- **Utilities** - The utilities are defined as

$$U_H(A_i^H, A_j^R, \mathcal{B}) = R(G_H^i) - C(\pi_{\Theta}^*(H)|\mathcal{B})$$

$$U_R(A_i^H, A_j^R, \mathcal{B}) = R(G_R^j) - C(\pi_{\Theta}^*(R)|\mathcal{B})$$

where symbols have their usual meaning.

As before, the Nash equilibriums in  $\text{GAPI-Bayesian}$  are given by action profiles  $\langle A_i^H, A_j^R \rangle$  such that the human has no reason to defect, i.e.  $\sum_{B \in \mathcal{B}} U_H(A_i^H, A_j^R, B)P(B) \geq \sum_{B \in \mathcal{B}} U_H(A_k^H : \forall k \neq i, A_j^R, B)P(B)$  while the robot also has no incentive to change, i.e.  $\sum_{B \in \mathcal{B}} U_R(A_i^H, A_j^R, B)P(B) \geq \sum_{B \in \mathcal{B}} U_R(A_i^H, A_k^R : \forall k \neq j, B)P(B)$ , given the distribution  $P$  over the beliefs  $\mathcal{B}$  of robot type. Similarly, the socially optimal solution is given by the action profiles  $\langle A_{i^*}^H, A_{j^*}^R \rangle$  where  $\{i^*, j^*\} = \arg \max_{i,j} \sum_{B \in \mathcal{B}} [U_H(A_i^H, A_j^R, B) + U_R(A_i^H, A_j^R, B)]P(B)$ .

## 5 Discussions and Future Work

The concept of Bayesian games lends  $\text{GAPI}$  to several interesting possibilities, and promising directions for future work, with respect to how interactions evolve with time.

### 5.1 Unrolling the Entire Game

Notice that we formulated the game such that each of the agents  $\phi$  has a set of goals  $\mathcal{G}^\phi$  to achieve. Thus  $\text{GAPI}$  immediately lends itself to a finite horizon dynamic game unrolled  $\max_{\phi \in \Phi} |\mathcal{G}^\phi|$  times, so that the agents can figure out their most effective long-term strategy and coalitions. Finding optimal policies in such cases will involve devising more

powerful approximations, and the ability to deal with issues such as synchronization and coalitions evolving across individual goal allocations. For  $\text{GAPI-Bayesian}$ , this also includes evolving beliefs as we will see below.

### 5.2 Impact of Intent Recognition

**Evolving Utilities.** Often, and certainly in the examples provided in Section 4.1, the behavior of the robot depends on understanding the intent(s) of its human counterpart. Thus the utilities will keep evolving based on the actions of the human after the goal has been selected. This is even more relevant in scenarios where communication is severely limited, when the agents in a coalition are not aware of the exact goals that the other agents have selected.

**Evolving Beliefs.** Intent recognition has a direct effect on the belief over the robot type itself. For example, as the human observes the actions of the robot, it can infer which behavior the robot is going to exhibit. Thus intent recognition over the robot’s actions will result in evolving belief of the human, as opposed to intent recognition over the human’s activities which informed the planning process and hence the utilities of the robot.

### 5.3 Implications of Implicit Preferences

Finally, as agents interact with each other over time, in different capacities as teammates and colleagues, their expectations over which agent is likely to form which form of coalition will also evolve. This will give the *prior* belief over the robot type that the human starts with, and will get updated as further interactions occur.

## 6 Conclusions

In conclusion, we introduced a two-player static game that can be used to form optimal coalitions on the go among two autonomous members of a human-robot society, with minimum prior coordination. We also looked at several properties of such games that may be used to make the problem tractable while still maintaining key properties of the game. Finally, we explored an extension of the game to a general Bayesian formulation when the human is not sure of the intent of the robot, and motivated the implications and expressiveness of this model. We believe the work will stimulate discussion on ad-hoc interaction among agents in the context of human-robot cohabitation settings and provide insight towards generating efficient synergy.

## Acknowledgments

This research is supported in part by the ONR grants N00014-13-1-0176, N00014-13-1-0519 and N00014-15-1-2027, and the ARO grant W911NF-13-1-0023. I would also like to give special thanks to Prof. Guoliang Xue (with the Department of Computer Science at Arizona State University) for his valuable support and inputs.

## References

- [Ackerman and Brânzei] Ackerman, M., and Brânzei, S. 2014. The authorship dilemma: Alphabetical or contribution? In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-agent Systems, AAMAS '14*, 1487–1488. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.
- [Chakraborti et al.] Chakraborti, T.; Briggs, G.; Talamadupula, K.; Zhang, Y.; Scheutz, M.; Smith, D.; and Kambhampati, S. 2015a. Planning for serendipity. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.
- [Chakraborti et al.] Chakraborti, T.; Zhang, Y.; Smith, D.; and Kambhampati, S. 2015b. Planning with stochastic resource profiles: An application to human-robot cohabitation. In *ICAPS Workshop on Planning and Robotics*.
- [Chakraborti et al.] Chakraborti, T.; Talamadupula, K.; Zhang, Y.; and Kambhampati, S. 2016. Interaction in human-robot societies. In *AAAI Workshop on Symbiotic Cognitive Systems*.
- [Mcdermott et al.] Mcdermott, D.; Ghallab, M.; Howe, A.; Knoblock, C.; Ram, A.; Veloso, M.; Weld, D.; and Wilkins, D. 1998. Pddl - the planning domain definition language. Technical Report TR-98-003, Yale Center for Computational Vision and Control.
- [Ray and Vohra] Ray, D., and Vohra, R. 2014. *Handbook of Game Theory*. Handbooks in economics. Elsevier Science.
- [Shoham and Tennenholtz] Shoham, Y., and Tennenholtz, M. 1992. On the synthesis of useful social laws for artificial agent societies. In *Proceedings of the Tenth National Conference on Artificial Intelligence, AAAI'92*, 276–281.
- [Talamadupula et al.] Talamadupula, K.; Briggs, G.; Chakraborti, T.; Scheutz, M.; and Kambhampati, S. 2014. Coordination in human-robot teams using mental modeling and plan recognition. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2957–2962.
- [Tambe] Tambe, M. 1997. Towards flexible teamwork. *J. Artif. Int. Res.* 7(1):83–124.
- [Zhang and Kambhampati] Zhang, Y., and Kambhampati, S. 2014. A formal analysis of required cooperation in multi-agent planning. In *ICAPS Workshop on Distributed Multi-Agent Planning (DMAP)*.
- [Zick and Elkind] Zick, Y., and Elkind, E. 2014. Arbitration and stability in cooperative games. *SIGecom Exch.* 12(2):36–41.
- [Zick, Chalkiadakis, and Elkind] Zick, Y.; Chalkiadakis, G.; and Elkind, E. 2012. Overlapping coalition formation games: Charting the tractability frontier. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems - Volume 2, AAMAS '12*, 787–794. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.