A Formal Framework for Studying Interaction in Human-Robot Societies

Tathagata Chakraborti[1] **Kartik Talamadupula**[2] **Yu Zhang**[1] **Subbarao Kambhampati**[1]

Department of Computer Science[1]
Arizona State University
Tempe, AZ 85281, USA
{tchakra2, yzhan442, rao}@asu.edu

Cognitive Learning Department[2]
IBM Thomas J. Watson Research Center
Yorktown Heights, NY 10598, USA
krtalamad@us.ibm.com

Abstract

As robots evolve into an integral part of the human ecosystem, humans and robots will be involved in a multitude of collaborative tasks that require complex coordination and cooperation. Indeed there has been extensive work in the robotics, planning as well as the human-robot interaction communities to understand and facilitate such seamless teaming. However, it has been argued that their increased participation as independent autonomous agents in hitherto human-habited environments has introduced many new challenges to the view of traditional human-robot teaming. When robots are deployed with independent and often self-sufficient tasks in a shared workspace, teams are often not formed explicitly and multiple teams cohabiting an environment interact more like colleagues rather than teammates. In this paper, we formalize these differences and analyze metrics to characterize autonomous behavior in such human-robot cohabitation settings.

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. We are now witnessing the development of autonomous agents that are especially designed to operate in predominantly human-inhabited environments often with completely independent tasks and goals. Examples of such agents include robotic security guards like Knightscope, virtual presence platforms like Double and iRobot Ava, and even autonomous assistance in hospitals such as Aethon TUG. Of particular fame are the CoBots (Rosenthal, Biswas, and Veloso 2010) that can ask for help from unknown humans, and thus interact with agents not directly involved in its plan. Indeed there has been a lot of work recently in the context of "human-aware" planning, both from a point of view of path planning (Sisbot et al. 2007; Kuderer et al. 2012) and task planning (Koeckemann, Pecora, and Karlsson 2014; Cirillo, Karlsson, and Saffiotti 2010), with the intention of making the robot's plans socially acceptable, e.g. resolving conflicts with the plans of fellow humans. Even though all of these scenarios involve significantly different levels of autonomy from the robotic agent, the underlying theme of autonomy in such settings involves the robot achieving some sense of independence of purpose in so much as its existence is not just defined by the goals of the humans around it but is rather contingent on tasks it is supposed to be achieving on its own. Thus the robots in a way become colleagues rather than teammates. This becomes even more prominent when we consider interactions between multiple independent teams in a human-robot cohabited environment. We thus postulate that the notions of coordination and cooperation between the humans and their robotic colleagues is inherently different from those investigated in existing literature on interaction in human-robot teams, and should rather reflect the kind of interaction we have come to expect from human colleagues themselves. Indeed recent work (Chakraborti et al. 2015a; Chakraborti et al. 2015b; Talamadupula et al. 2014) hints at these distinctions, but has neither made any attempt at formalizing these ideas, nor provided methods to quantify behavior is such settings. To this end, we propose a formal framework for studying inter-team and intra-team interactions in human-robot societies, show how existing metrics are grounded in this framework and propose newer metrics that are useful for evaluating performance of autonomous agents in such environments.

1 Human Robot Cohabitation

At some abstracted level, agents in any environment can be seen as part of a team achieving a high level goal. Consider, for example, your university or organization. At a micro level, it consists of many individual labs or groups that work independently on their specific tasks. But when taken as a whole, the entire institute is a team trying to achieve some higher order tasks like increasing its relative standing among its peers or competitors. So in the discussion that follows, we talk about environments, and teams or colleagues acting within it, in the context of the goals they achieve.

1.1 Goal-oriented Environments

Definition 1.0 *A goal-oriented 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 a set of objects in the environment, $\Phi \subseteq \mathbb{O}$ is the set of agents, $\mathcal{G} = \{g \mid g \subseteq \mathbb{F}_{\mathbb{O}}\}$ is the set of goals that these agents are tasked with, and $\Lambda \subseteq \mathbb{O}$ is the set of resources that are required by the agents to achieve their goals. Each goal has* *a reward $R(g) \in \mathbb{R}^+$ associated with it.*[1]

These agents and goals are, of course, related to each other by their tasks, and these relationships determine the nature of their interactions in the environment, i.e. in the form of teams or colleagues. Before we formalize such relations, however, we would look at the way the agent models are defined. We use PDDL (Mcdermott et al. 1998) models for the rest of the discussion, as described below, but most of the discussion easily generalizes to other modes of representation. The domain model D_ϕ of an agent $\phi \in \Phi$ is defined as $D_\phi = \langle \mathbb{F}_\mathbb{O}, A_\phi \rangle$, where A_ϕ 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, $\mathbb{P}_a \subseteq \mathbb{F}_\mathbb{O}$ is the list of pre-conditions that must hold for the action a to be applicable in a particular state $S \subseteq \mathbb{F}_\mathbb{O}$ of the environment; and $\mathbb{E}_a = \langle eff^+(a), eff^-(a) \rangle$, $eff^\pm(a) \subseteq \mathbb{F}_\mathbb{O}$ 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 $\delta(a, S) = (S \backslash eff^-(a)) \cup eff^+(a)$ if $\mathbb{P}_a \subseteq S$; $\perp$ otherwise.

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

We will now introduce the concept of a super-agent transformation 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 able to do. This will help us later to formalize the nature of interactions among agents.

Definition 1.1a *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_θ is the transformation from the individual domain models to a composite domain model given by $D_\theta = \langle \mathbb{F}_\mathbb{O}, \bigcup_{\phi \in \theta} A_\phi \rangle$.*

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 1.1b *The planning problem of a super-agent Θ is similarly given by $\Pi_\Theta = \langle \mathbb{F}, \mathbb{O}, D_\theta, \mathbb{I}_\theta, \mathbb{G}_\theta \rangle$ 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 $\pi_\theta = \langle \mu_1, \mu_2, \ldots, \mu_{|\pi_\theta|} \rangle$ where $\mu_i = \{a_1, \ldots, a_{|\theta|}\}$, $\mu(\phi) = a \in A_\phi \; \forall \mu \in \pi_\theta$ such that $\delta'(\mathbb{I}_\theta, \pi_\theta) \models \mathbb{G}_\theta$, where the modified transition function $\delta'(\mu, s) = (s \setminus \bigcup_{a \in \mu} \text{eff}^-(a)) \cup \bigcup_{a \in \mu} \text{eff}^+(a)$. We denote the set of all such plans as π_Θ.*

The cost of a composite plan is $C(\pi_\theta) = \sum_{\mu \in \pi_\theta} \sum_{a \in \mu} \mathbb{C}_a$ and π^*_θ is optimal if $\delta'(\mathbb{I}_\theta, \pi_\theta) \models \mathbb{G}_\theta \implies C(\pi^*_\theta) \leq C(\pi_\theta)$. The composite plan can thus be viewed as a union of plans contributed by each agent $\phi \in \theta$ so that ϕ's component can be written as $\pi_\theta(\phi) = \langle a_1, a_2, \ldots, a_n \rangle$, $a_i = \mu_i(\phi) \; \forall \mu_i \in \pi_\Theta$. Now we will define the relations among the components of the environment $\mathcal{E}$ in terms of these agent models.

Definition 1.2 *At any given state $S \subseteq \mathbb{F}_\mathbb{O}$ of the environment $\mathcal{E}$, a goal-agent correspondence is defined as the relation $\tau : \mathcal{G} \rightarrow \mathcal{P}(\Phi)$; $\mathcal{G}, \Phi \in \mathcal{E}$, that induces a set of super-agents $\tau(g) = \{\Theta \mid \Pi_\Theta = \langle \mathbb{F}, \mathbb{O}, D_\theta, S, g \rangle$* has a solution, i.e. $\exists \pi$ *s.t.* $\delta(\pi, S) \models g\}$.

In other words, $\tau(g)$ gives a list of sets of agents in the environment that are capable of performing a specific task g. We will see in the next section how the notions of teammates and colleagues are derived from it.

1.2 Teams and Colleagues

Definition 2.0 *A team T_g w.r.t. a goal $g \in \mathcal{G}$ is defined as any super-agent $\Theta = \langle \theta, D_\theta \rangle \in \tau(g)$ iff $\nexists \phi \in \theta$ such that $\Theta' = \langle \theta \setminus \phi, D_{\theta \setminus \phi} \rangle$ and $\pi_\Theta = \pi_{\Theta'}$.*

This means that any super-agent belonging to a particular goal-agent correspondence defines a team w.r.t that specific goal when every agent that forms the super-agent plays *some* part in the plans that achieves the task described by g, i.e. the super-agent cannot use the same plans to achieve g if an agent is removed from its composition. This, then, leads to the concept of *strong*, *weak*, or *optimal* teams, depending on if the composition of the super-agent is *necessary*, *sufficient* or *optimal* respectively (note that an optimal team may or may not be a strong team).

Definition 2.0a *A team $T^s_g = \langle \theta, D_\theta \rangle \in \tau(g)$ w.r.t a goal $g \in \mathcal{G}$ is strong iff $\nexists \phi \in \theta$ such that $\langle \theta \setminus \phi, D_{\theta \setminus \phi} \rangle \in \tau(g)$. A team T^w_g is weak otherwise.*

Definition 2.0b *A team $T^o_g = \langle \theta, D_\theta \rangle \in \tau(g)$ w.r.t a goal $g \in \mathcal{G}$ is optimal iff $\forall \Theta' \in \tau(g), C(\pi^*_\theta) \leq C(\pi^*_{\theta'})$.*

This 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). In this paper, we are more concerned about the *consequences* of such team formations on teaming metrics. So, with these different types of teams we have seen thus far, the question we ask is: *What is the relation among the rest of the agents in the environment? How do these different teams interact among and between themselves?*

Definition 2.1 *The set of teams in $\mathcal{E}$ are defined by the relation $\kappa : \mathcal{G} \rightarrow R(\tau)$; $\mathcal{G} \in \mathcal{E}$, where $\kappa(g) \in \tau(g)$ denotes the team assigned to the goal g, $\forall g \in \mathcal{G}$.*

This, then, gives rise to the idea of collegiality among agents, due to both inter-team and intra-team interactions. Note that how useful or necessary such interactions are will depend on whether the colleagues can contribute to each other's goals, or to what extent they influence their respective plans, which leads us to the following two definitions of colleagues based on the concept of teams.

Definition 2.2a *Let $\kappa(g) = \langle \theta_1, D_{\theta_1} \rangle, \kappa(g') = \langle \theta_2, D_{\theta_2} \rangle$ be two teams in $\mathcal{E}$. An agent $\phi_1 \in \theta_1$ is a type-1 colleague to an agent $\phi_2 \in \theta_2$ when $\kappa'(g) = \langle \theta_1 \cup \phi_1, D_{\theta_1 \cup \phi_1} \rangle$ is a weak team w.r.t. the goal g.*

[1] $S_\mathbb{O}$ is $S \subseteq \mathbb{F}$ instantiated or grounded with objects from $\mathbb{O}$.

Definition 2.2b *Agents $\phi_1, \phi_2 \in \Phi$ are type-2 colleagues when $\forall \kappa(g) = \langle \theta, D_\theta \rangle$ s.t. $\{\phi_1, \phi_2\} \cap \theta \neq \varnothing, \{\phi_1, \phi_2\} \notin \theta \wedge \kappa'(g) = \langle \theta \cup \{\phi_1, \phi_2\}, D_{\theta \cup \{\phi_1, \phi_2\}} \rangle$ is a weak team.*

Thus type-1 colleagues can potentially contribute to the plans of their colleagues, while type-2 colleagues cannot. Plans of type-2 colleagues can, however, influence each other (for example due to conflicts on usage of shared resources), while type1-colleagues are capable of becoming teammates dynamically during plan execution.

Humans in the loop. Instead of a general set of agents, we define the set of agents θ in a super-agent as composition of humans and robots $\theta = h(\theta) \cup r(\theta)$ so that the domain model of the super-agent is also a composition of the human and robot capabilities $D_\theta = \bigcup_{\phi \in h(\theta)} \bigcup_{\phi \in r(\theta)} A_\phi = h(D_\theta) \cup r(D_\theta)$. We denote the communication actions of the super-agent as the subset $c(D_\theta) \subseteq D_\theta$.

2 Metrics for Human Robot Interaction

2.1 Metrics for Human Robot Teams

We will now ground popular (Olsen Jr. and Goodrich 2003; Steinfeld et al. 2006; Hoffman and Breazeal 2007; Hoffman 2013) metrics for human-robot teams in our current formulation.

Task Effectiveness These are the metrics that measure the effectiveness of a team in completing its tasks.

- **Cost-based Metrics** - This simply measures the cost $\sum_{g \in \kappa^{-1}(\Theta)} C(\pi^*_\Theta)$ of all the (optimal) plans a specific team executes (for all the goals it has been assigned to).
- **Net Benefit Based Metrics** - This is based on both plan costs as well as the value of goals and is given by $\sum_{g \in \kappa^{-1}(\Theta)} R(g) - C(\pi^*_\Theta)$.
- **Coverage Metrics** - Coverage metrics for a particular team determine the diversity of its capabilities in terms of the number of goals it can achieve $|\kappa^{-1}(\Theta)|$.

Team Effectiveness These measure the effectiveness of (particularly human-robot) teaming in terms of communication overhead and smoothness of coordination.

- **Neglect Tolerance** - This measures how long the robots in a team Θ is able to perform well without human intervention. We can measure this as $NT = \max\{|i - j| \text{ s.t. } h(D_\theta) \bigcap_{\phi \in \theta} \pi^*_\Theta(\phi)[i : j] = \varnothing\}$.
- **Interaction Time** - This is given by $IT = \sum |\{i \mid c(D_\theta) \cap \pi^*_\Theta[i] \neq \varnothing\}|$, and measures the time spent by a team Θ in communication.
- **Robot Attention Demand** - Measures how much attention the robot is demanding and is given by $\frac{IT}{IT+NT}$.
- **Secondary Task Time** - This measures the "distraction" to a team, and can be expressed as time not spent on achieving a given goal g, i.e. $STT = |\{i \mid \pi^*_\Theta[i] := \varnothing \wedge \delta'(s, \pi^*_\Theta) \models g\}|$.
- **Free Time** - $FT = 1 - RAD$ is a measure of the fraction of time the humans are not interacting with the robot.
- **Human Attention Demand** - $HAD = FT - h(STT)$ where $h(STT) = |\{i \mid \pi^*_\Theta[i] \cap h(D_\theta) := \varnothing \wedge \delta'(s, \pi^*_\Theta) \models g\}| / |\pi^*_\Theta|$ is the time humans spend on the secondary task.
- **Fan Out** - This is a measure of the communication load on the humans, and consequently the number of robots that should participate in a human-robot team, and is proportional to $FO \propto |h(\theta)|/RAD$.
- **Interaction Time** - Measures how quickly and effectively interaction takes places as $IT = \frac{NT(1-STT)}{STT}$.
- **Robot Idle Time** - Captures inconsistency or irregularity in coordination from the point of view of the robotic agent, and can be measured as the amount of time the robots are idle, i.e. $RIT = |\{i \mid r(D_\theta) \cap \pi^*_\Theta[i] = \varnothing|$.
- **Concurrent Activity** - We can talk of concurrency within a team as the time that humans and robots are working concurrently $CA_1 = |\{i \mid r(D_\theta) \cap h(D_\theta) \cap \pi^*_\Theta[i] \neq \varnothing\}|$ and also across teams as the maximum time teams are operating concurrently $CA_2 = \max\{|\{i \mid \pi_\Theta[i] \neq \varnothing \wedge \pi_{\Theta'}[i] \neq \varnothing\}| \, \forall \Theta, \Theta' \in \mathcal{R}(\kappa)\}$.

In measuring performance of agents in cohabitation, both as teammates and colleagues, we would still like to reduce interactions times and attentions demand, while simultaneously increasing neglect tolerance and concurrency. However, as we will see in Section 3, these metrics do not effectively capture all the implications of the interactions desired in human-robot cohabitation. So the purpose of the rest of our paper is to establish metrics that can measure the effective behavior of human-robot colleagues, and to see to what extent they can capture desired behaviors of robotic colleagues suggested in existing literature.

2.2 Metrics for Human Robot Colleagues

We will now propose new metrics that are useful for measuring collegial interactions, see how they differ from teaming metrics discussed so far, and then relate them to existing work on human-robot cohabitation.

Task Effectiveness The measures for task effectiveness must take into account that agents are not necessarily involved in their assigned team task only.

- **Altruism** - This is a measure of how useful it is for a robotic agent r to showcase altruistic behavior in assisting their human colleagues, and is given by the ratio of the gain in utility by adding a robotic colleague to a team Θ to the decrease in utility of plans of the teams r is involved in $|\pi^*_\Theta - \pi^*_{\langle \theta \cup r, D_{\theta \cup r} \rangle}| / \sum_{\Theta = \langle \theta, D_\theta \rangle \text{ s.t. } r \in \theta} \Delta|\pi^*_\Theta|$. For such a dynamic coalition to be useful, r must be a type-1 colleague to the agents $\theta \in \Theta$.
- **Lateral Coverage** - This measures how deviating from optimal team compositions can achieve global good in terms of number of goals achieved by a team, $LT = \sum_{T_g = \kappa(g), \forall g \in \mathcal{G}} \{[|\kappa^{-1}(T_g)| - |\kappa^{-1}(T^o_g)|] / |\kappa^{-1}(T^o_g)|\}$ across all the teams that have been formed in $\mathcal{E}$.
- **Social Good** - Many times, while planning with humans in the loop, cost optimal plans are not necessarily the optimal plans in the social context. This is useful to measure particularly when agents are interacting outside teams, and the compromise in team utility is compensated by the gain in mutual utility of colleagues. This can be expressed as $\sum_{g \in \mathcal{G}} \{C(\pi_{\kappa(g)}) - C(\pi^*_{\kappa(g)})\}$.

Interaction Effectiveness The team effectiveness measures need to be augmented with measures corresponding to interactions among non-team members. While all these metrics are relevant for robotic colleagues as well, they become particularly important in human-robot interactions, where information is often not readily sharable due to higher cognitive mismatch, so as to reduce cognitive demand/overload.

- **Interaction Time** - In addition to *Interaction Time* for human-robot teams, and measures derived from it, we propose two separate components of interaction time for general human-robot cohabitation scenarios.
  - **External Interaction Time** - This is the time spent by agents interacting with type-1 colleagues (EIT_1).
  - **Extraneous Interaction Time** - This is the time spent by agents interacting with type-2 colleagues (EIT_2).
- **Compliance** - This refers to how much actions of an agent disambiguate its intentions. Though relevant for both, this becomes even more important in absence of teams, when information pertaining to goals or plans are not necessarily sharable. Thus the intention should be to maximize the probability $P(\mathbb{G}_\theta = g \mid s = \delta(\pi_\theta[1:i], \mathbb{I}_\theta)),\ \kappa(g) = \langle \theta, D_\theta \rangle, \forall g \in \mathcal{G}$ given any stage i of plan execution and $P(\cdot)$ is a generic goal recognition algorithm. This can be relevant both in terms of disambiguating goals (Keren, Gal, and Karpas 2014) or explaining plans given a goal (Zhang, Zhuo, and Kambhampati 2015).
- **External Failure** - This is the number of times optimal plans fail when resources are contested among colleagues.
- **Stability** - Of course with continuous interactions, team formations change, so this gives a measure of stability of the system as a whole. If teams $\kappa(g) = \langle \theta_1, D_{\theta_1} \rangle$ and $\kappa(g) = \langle \theta_2, D_{\theta_2} \rangle$ achieves a goal $g \in \mathcal{G}$ at two different instances, then stability $S = \sum_{g \in \mathcal{G}} |\theta_1 \cap \theta_2| / |\theta_1||\theta_2|$.

3 Discussion and Related Work

We will now investigate the usefulness of the proposed metrics in quantifying behavioral traits proposed in existing literature as desirable among cohabiting human and robots.

Human-Aware Planning. In (Koeckemann, Pecora, and Karlsson 2014; Cirillo, Karlsson, and Saffiotti 2010) the authors talk of adapting robot plans to suit social norms (e.g. not to vacuum a room while a human is asleep). Clearly, this involves the robots departing from their preferred plans to conform to human preferences. In such cases, involving assistive robots, measures of *Altruism* and *Social Good* become particularly relevant, while it is also crucial to reduce unwanted interactions ($EIT_1 + EIT_2$).

Planning with Resource Conflicts. In (Chakraborti et al. 2015b) the authors outline an approach for robots sharing resources with humans to compute plans that minimize conflicts in resource usage. Thus, this line of work is aimed at reducing *External Failures*, while simultaneously increasing *Social Good*. Measures of *Stability* and *Compliance* become relevant, to capture evolving beliefs and their consequences on plans. *Extraneous Interaction Time* is also an important measure, since additional communication is always a proxy to minimizing coordination problems between colleagues.

Planning for Serendipity. In (Chakraborti et al. 2015a) the authors propose a formulation for the robot to produce positive exogenous events during the execution of the human's plans, i.e. interventions which will be useful to the human regardless of whether he was expecting assistance from the robot. This work particularly looks at planning for *Altruism*. Increasing *Compliance* in agent behavior can provide better performance in this regard. Further, *External Interaction* is crucial in such cases for forming such impromptu coalitions among colleagues.

Relation to Metrics in Human Factor Studies It is useful to see an example of how the general formulation of metrics we discussed so far are actually grounded in human factors studies (Zhang et al. 2015) of scenarios that display some aspects of collegial interaction. The environment studied was a disaster response scenario, involving an autonomous robot that may or may not chose to proactively help the human. The authors used *External Interaction Time* or EIT_1 to measure the effectiveness of proactive support (how often the proactive support resulted in further deliberation over goals), while *Lateral Coverage* (in terms of number of people rescued) showed the effectiveness of proactive support. Further, qualitative analysis on acceptance and usefulness of agents that display proactive support are closely related to measures such as *Social Good* and *Altruism*.

Work on Ad-hoc Coalition Formations Given the framework we have discussed thus far, the question is then, apart from *measuring* performance, how we can use it to *facilitate* collegial interactions among agents. Especially relevant in such scenarios are work on ad-hoc coalition formation among agents sharing an environment but not necessarily goals (Stone et al. 2010). In (Chakraborti et al. 2016) we show how this framework may be used to cut down on prior coordination while forming coalitions.

4 Conclusion and Future Work

In conclusion, we discussed interaction in human-robot societies involving multiple teams of humans and robots in the capacity of teammates or as colleagues, provided a formal framework for talking about various modes of cooperation, and reviewed existing metrics and proposed new ones that can capture these different modalities of teaming or collegial behavior. Finally we discussed how such metrics can be useful in evaluating existing works in human-robot cohabitation. One line of future inquiry would be to see how such quantitative metrics are complemented by qualitative feedback from human factor studies, to establish what the desired trade-offs are, in order to ensure well-informed design of symbiotic systems involving humans and robots.

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.

References

[Aethon TUG] Aethon TUG. Intralogistics automation platform for hospitals. http://www.aethon.com/.

[Chakraborti et al. 2015a] Chakraborti, T.; Briggs, G.; Talamadupula, K.; Zhang, Y.; Scheutz, M.; Smith, D.; and Kambhampati, S. 2015a. Planning for serendipity. In *International Conference on Intelligent Robots and Systems*.

[Chakraborti et al. 2015b] 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. 2016] Chakraborti, T.; Dondeti, V.; Meduri, V. V.; and Kambhampati, S. 2016. A game theoretic approach to ad-hoc coalition formation in human-robot societies. In *AAAI Workshop on Multi-Agent Interaction without Prior Coordination*.

[Cirillo, Karlsson, and Saffiotti 2010] Cirillo, M.; Karlsson, L.; and Saffiotti, A. 2010. Human-aware task planning: An application to mobile robots. *ACM Trans. Intell. Syst. Technol.* 1(2):15:1–15:26.

[Double] Double. The ultimate tool for telecommuting. http://www.doublerobotics.com/.

[Hoffman and Breazeal 2007] Hoffman, G., and Breazeal, C. 2007. Effects of anticipatory action on human-robot teamwork: Efficiency, fluency, and perception of team. In *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on*, 1–8.

[Hoffman 2013] Hoffman, G. 2013. Evaluating fluency in human-robot collaboration. In *Robotics: Science and Systems (RSS) Workshop on Human-Robot Collaboration*.

[iRobot Ava] iRobot Ava. Video collaboration robot. http://www.irobot.com/For-Business.aspx.

[Keren, Gal, and Karpas 2014] Keren, S.; Gal, A.; and Karpas, E. 2014. Goal recognition design. In *Proceedings of the Twenty-Fourth International Conference on Automated Planning and Scheduling, ICAPS 2014, Portsmouth, New Hampshire, USA, June 21-26, 2014*.

[Knightscope] Knightscope. Autonomous data machines. http://knightscope.com/about.html.

[Koeckemann, Pecora, and Karlsson 2014] Koeckemann, U.; Pecora, F.; and Karlsson, L. 2014. Grandpa hates robots - interaction constraints for planning in inhabited environments. In *Proc. AAAI-2010*.

[Kuderer et al. 2012] Kuderer, M.; Kretzschmar, H.; Sprunk, C.; and Burgard, W. 2012. Feature-based prediction of trajectories for socially compliant navigation. In *Proceedings of Robotics: Science and Systems*.

[Mcdermott et al. 1998] 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,.

[Olsen Jr. and Goodrich 2003] Olsen Jr., D., and Goodrich, M. A. 2003. Metrics for evaluating human-robot interactions. In *Performance Metrics for Intelligent Systems*.

[Rosenthal, Biswas, and Veloso 2010] Rosenthal, S.; Biswas, J.; and Veloso, M. 2010. An effective personal mobile robot agent through symbiotic human-robot interaction. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Volume 1 - Volume 1*, AAMAS '10, 915–922. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

[Shoham and Tennenholtz 1992] 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. AAAI Press.

[Sisbot et al. 2007] Sisbot, E.; Marin-Urias, L.; Alami, R.; and Simeon, T. 2007. A human aware mobile robot motion planner. *Robotics, IEEE Transactions on* 23(5):874–883.

[Steinfeld et al. 2006] Steinfeld, A.; Fong, T.; Kaber, D.; Lewis, M.; Scholtz, J.; Schultz, A.; and Goodrich, M. 2006. Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction*, 33–40.

[Stone et al. 2010] Stone, P.; Kaminka, G. A.; Kraus, S.; and Rosenschein, J. S. 2010. Ad hoc autonomous agent teams: Collaboration without pre-coordination. In *Proceedings of the Twenty-Fourth Conference on Artificial Intelligence*.

[Talamadupula et al. 2014] 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 1997] Tambe, M. 1997. Towards flexible teamwork. *J. Artif. Int. Res.* 7(1):83–124.

[Zhang and Kambhampati 2014] 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)*.

[Zhang et al. 2015] Zhang, Y.; Narayanan, V.; Chakraborti, T.; and Kambhampati, S. 2015. A human factors analysis of proactive support in human-robot teaming. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*.

[Zhang, Zhuo, and Kambhampati 2015] Zhang, Y.; Zhuo, H. H.; and Kambhampati, S. 2015. Plan explainability and predictability for cobots. *CoRR* abs/1511.08158.