Coalitional Bargaining with Agent Type Uncertainty

Georgios Chalkiadakis and **Craig Boutilier**
Department of Computer Science
University of Toronto, Toronto, Canada
{ gehalk, cebly }@cs.toronto.edu

Abstract

Coalition formation is a problem of great interest in AI, allowing groups of autonomous, individually rational agents to form stable teams. Automating the negotiations underlying coalition formation is, naturally, of special concern. However, research to date in both AI and economics has largely ignored the potential presence of uncertainty in coalitional bargaining. We present a model of discounted coalitional bargaining where agents are uncertain about the types (or capabilities) of potential partners, and hence the value of a coalition. We cast the problem as a Bayesian game in extensive form, and describe its Perfect Bayesian Equilibria as the solutions to a polynomial program. We then present a heuristic algorithm using iterative coalition formation to approximate the optimal solution, and evaluate its performance.

1 Introduction

Coalition formation, widely studied in game theory and economics [8], has attracted much attention in AI as means of dynamically forming partnerships or teams of cooperating agents. While most models of coalition formation (e.g., *coalitional bargaining processes*) assume that agents have full knowledge of *types* of their potential partners, in most natural settings this will not be the case. Generally, agents will be uncertain about various characteristics of others (e.g., their capabilities), which in turn imposes uncertainty on the value of any coalition. This presents the opportunity to learn about the types of others based on their behavior during negotiation and by observing their performance in settings where coalitions form repeatedly. Agents must be able to form coalitions and divide the generated value even in such settings.

Here we present a model of discounted coalitional bargaining under agent type uncertainty. We formulate this as a Bayesian extensive game with observable actions [8], where the actions correspond to proposing choices of potential partners and a payoff allocation, or accepting or rejecting such proposals. Our model generalizes related bargaining models by explicitly dealing with uncertainty about agent types (or capabilities) and coalitional values. We formulate the perfect Bayesian equilibrium (PBE) solution of this game as a decidable polynomial program. The complexity of the program makes it intractable for all but trivial problems, so we propose an alternative heuristic algorithm to find good agent strategies in the coalitional bargaining game. Preliminary experiments illustrate the performance of this heuristic approach.

Although there is a considerable body of work on coalitional bargaining, no existing models deal with explicit type uncertainty. Okada [7] suggests a form of coalitional bargaining where agreement can be reached in one bargaining round if the proposer is chosen *randomly*. Chatterjee et al. [3] present a bargaining model with a fixed proposer order, which results in a delay of agreement. Neither model deals with type uncertainty—instead, they focus on calculating subgame-perfect equilibria (SPE). Suijs et al. [9] introduce *stochastic cooperative games (SCGs)*, comprising a set of agents, a set of coalitional actions, and a function assigning to each action a random variable with finite expectation, representing action-dependent coalition payoff. Though they provide strong theoretical foundations for games with this restricted form of action uncertainty, they do not model explicitly a coalition formation process. Kraus et al. [4] model coalition formation under a restricted form of uncertainty regarding coalitional values in a *request for proposal* domain. However, type uncertainty is not captured; rather, the mean value of coalitions is *common knowledge*, and a "manager" handles proposals (they also focus on social welfare maximization rather than individual rationality).

Chalkiadakis and Boutilier [2] propose an explicit model of type uncertainty and show how this translates into coalitional value uncertainty. We adopt their model in our paper. However, their results focus on stability concepts and how coalitions evolve during repeated interaction, as agents gradually learn more about each other's capabilities (in reinforcement learning style). The actual coalition formation processes used are fairly simple and are not influenced by strategic considerations, nor do agents update their beliefs about other agents' types during bargaining. Our work analyzes the actual bargaining process in more depth.

2 Bayesian Coalitional Bargaining

We begin by describing the Bayesian coalition formation model and then define our coalitional bargaining game.

We assume a set of agents $N = \{1, \ldots, n\}$, and for each agent i a finite set of possible *types* T_i. Each agent i has a specific type $t \in T_i$. We let $T = \times_{i \in N} T_i$ denote the set of type profiles. Each i knows its own type t_i, but not those of other agents. Agent i's *beliefs* μ_i comprise a joint distribution

over T_{-i}, where $\mu_i(t_{-i})$ is the probability i assigns to other agents having type profile t_{-i}. Intuitively, i's type reflects its "abilities;" and its beliefs about the types of others capture its uncertainty about their abilities. For instance, if a carpenter wants to find a plumber and electrician with whom to build a house, her decision to propose (or join) such a partnership, to engage in a specific type of project, and to accept a specific share of the surplus generated should all depend on her probabilistic assessment of their abilities.

A *coalition* $C \subseteq N$ of members with actual types t_C has a value $V(t_C)$, representing the value this group can achieve by acting optimally. However, this simple *characteristic function* representation of the model [8] is insufficient, since this value is not common knowledge. An agent i can only assess the *expected value* of such a coalition based on its beliefs: $V_i(C) = \sum_{t_C \in T_C} \mu_i(t_C)V(t_C)$.

A coalition structure CS partitions N into coalitions of agents. A *payoff allocation* $P = \langle x_i \rangle$, given the stochastic nature of payoffs in this setting, assigns to each agent in coalition C its *share* of the value attained by C (and must be such that $\sum_{i \in C} x_i = 1$ for each $C \in CS$). Chalkiadakis and Boutilier [2] define the *Bayesian core* as a generalization of the standard core concept, capturing an intuitive notion of stability in the Bayesian coalition formation game.

While coalition structures and allocations can sometimes be computed centrally, in many situations they emerge as the result of some bargaining process among the agents, who propose, accept and reject partnership agreements [3]. We now define *a (Bayesian) coalitional bargaining game* for the model above as a *Bayesian extensive game with observable actions*. The game proceeds in stages, with a randomly chosen agent proposing a coalition and allocation of payments to partners, who then accept or reject the proposal.

A finite set of *bargaining actions* is available to the agents. A bargaining action corresponds to either some *proposal* $\pi = \langle C, P_C \rangle$ to form a coalition C with a specific payoff allocation P_C specifying payoff shares x_i to each $i \in C$, or to the acceptance or rejection of such a proposal. The finite-horizon game proceeds in S stages, and initially all agents are *active*. At the beginning of stage $s \leq S$, one of the (say n) active agents i is chosen randomly with probability $\gamma = \frac{1}{n}$ to make a proposal $\langle C, P_C \rangle$ (with $i \in C$). Each other $j \in C$ *simultaneously* (without knowledge of other responses) either accepts or rejects this proposal. If all $j \in C$ accept, the agents in C are made inactive and removed from the game. Value $V_s(t_C) = \delta^{s-1}V(t_C)$ is realized by C at s, and split according to P_C, where $\delta \in (0, 1)$ is the discount factor.[1] If any $j \in C$ rejects the proposal, the agents remain active (no coalition is formed). At the end of a stage, the responses are observed by all participants. At the end of stage S, any i not in any coalition receives its discounted reservation value $\delta^{S-1}V(t_i)$ (discounted singleton coalition value).

3 Perfect Bayesian Equilibrium

The coalitional bargaining game described above is clearly an extensive form Bayesian game. We assume each agent will adopt a suitable *behavioral strategy*, associating with each node in the game tree at which it must make a decision a distribution over action choices for *each of its possible types*. Furthermore, since it is uncertain about the types of other agents, its observed history of other agents' proposals and responses give it information about their types (assuming they are rational). Thus, the preferred solution concept is that of a *perfect Bayesian equilibrium (PBE)* [8]. A PBE comprises a profile of behavioral strategies for each agent as well a *system of beliefs* dictating what each agent believes about the types of its counterparts at each node in the game tree. The standard rationality requirements must also hold: the strategy for each agent maximizes its expected utility given its beliefs; and each agent's beliefs are updated from stage to stage using Bayes rule, given the specific strategies being played. In this section, we formulate the constraints that must hold on both strategies and beliefs in order to form a PBE.

Let σ_i denote a behavioral strategy for i, mapping information sets (or observable histories h) in the game tree at which i must act into distributions over admissible actions $A(h)$. If i is a proposer at h (at stage s), let $A(h) = \mathcal{P}$, the finite set of proposals available at h. Then $\sigma_i^{h,t_i}(\pi)$ denotes the (behavioral strategy) probability that i makes proposal $\pi \in \mathcal{P}$ at h given its type is t_i. If i is a responder at h, then $\sigma_i^{h,t_i}(y)$ is the probability with which i accepts the proposal on the table (says *yes*) at h (and $\sigma_i^{h,t_i}(n) = 1 - \sigma_i^{h,t_i}(y)$ is the probability i says *no*). Let μ_i denote i's beliefs with $\mu_i^{h,t_i}(t_{-i})$ being i's beliefs about the types of others at h given its own type is t_i.

We define the PBE constraints for the game by first defining the values to (generic) agent i at each node and information set in the game tree, given a fixed strategy for other agents, and the rationality constraints on his strategies and beliefs. We proceed in stages.

(1) Let ξ be a proposal node for i at history h at stage s. Since the only uncertainty in information set h involves the types of other agents, each $\xi \in h$ corresponds to one such type vector $t_{-i} \in T_{-i}$; let $h(t_{-i})$ denote this node in h. The value to i of a proposal $\pi = \langle C, P_C \rangle$ at $h(t_{-i})$ is:

$$q_i^{h(t_{-i}),t_i}(\pi) = p_{acc}^{h(t_{-i})}(\pi)x_i V_s(t_C) + \sum_r p^{h(t_{-i})}(\pi, r)q_i^{\xi/\pi/r,t_i}$$

where: $p_{acc}^{h(t_{-i})}(\pi)$ is the probability that all $j \in C$ (other than i) accept π (this is easily defined in terms of their fixed strategies); x_i is i's payoff share in P_C; r ranges over response vectors in which at least one $j \in C$ refuses the proposal; $p^{h(t_{-i})}(\pi, r)$ denotes the probability of such a response; and $q_i^{\xi/\pi/r,t_i}$ denotes the *continuation payoff* for i at stage $s+1$ at the node $\xi/\pi/r$ (following n after proposal π and responses r). This continuation payoff is defined (recursively) below. The value of π at history h (as opposed to a node) is determined by taking the expectation w.r.t. possible types: $q_i^{h,t_i}(\pi) = \sum_{t_{-i}} \mu_i^{h,t_i}(t_{-i})q_i^{h(t_{-i}),t_i}(\pi)$.

(2) Suppose i is a responder at node $\xi = h(t_{-i})$ in history h at stage s. As above, ξ corresponds to specific t_{-i} in h. W.l.o.g. we can assume i is the first responder (since all responses are simultaneous). Let $p_{acc}^{h(t_{-i})}(\pi)$ denote the probability that all other responders accept π. We then define the

[1] Agents could have different δ's. As long as these are common knowledge, our analysis holds with only trivial modifications.

value to i of accepting π at ξ as:

$$q_i^{h(t_{-i}),t_i}(y) = p_{acc}^{h(t_{-i})}(\pi)x_i V_s(t_C) + \sum_r p^{h(t_{-i})}(\pi, r) q_i^{\xi/y/r,t_i}$$

where again r ranges over response vectors in which at least one $j \in C$, $j \neq i$, refuses π; $p^{h(t_{-i})}(\pi, r)$ is the probability of such a response; and $q_i^{\xi/y/r,t_i}$ is the continuation payoff for i at stage $s+1$ after responses r by its counterparts. The value of accepting at h is given by the expectation over type vectors t_C w.r.t. i's beliefs μ_i^{h,t_i} as above.

The value of rejecting π at $\xi = h(t_{-i})$ is the expected continuation payoff at stage $s+1$:

$$q_i^{h(t_{-i}),t_i}(n) = \sum_r p^{h(t_{-i})}(\pi, r) q_i^{\xi/n/r,t_i}$$

(where r ranges over all responses, including pure positive responses, of the others).

(3) We have defined the value for i taking a specific action at any of its information sets. It is now straightforward to define the value to i of reaching any other stage s node controlled by $j \neq i$ or by nature (i.e., chance nodes where a random proposer is chosen).

First we note that, by assuming i responds "first" to any proposal, our definition above means that we need not compute the value to i at any response node (or information set) controlled by j. For an information set h_j where j makes a proposal, consider a node $\xi = h_j(t_j)$ where j is assumed to be of type t_j. Then, j's strategy $\sigma_j^{h_j,t_j}$ specifies a distribution over proposals π (determined given the values $q_j^{h_j,t_j}(\pi)$ which can be calculated as above, and j's type t_j). Agent i's value $q_i^{t_i,h_j(t_j)}$ at this node is given by the expectation (w.r.t. this strategy distribution) of its accept or reject values (or if it is not involved in a proposal, its expected continuation value at stage $s+1$ given the responses of others). Its value at h_j is then $Q_i^{t_i}(h_j) = \sum_{t_j} \mu_i^{h_j,t_i}(t_j) q_i^{t_i,h_j(t_j)}$. We define $Q_i^{t_i}(h_i)$ (where i is the proposer) as in Case 1 above.

Finally, i's value at information set h that defines the stage s continuation game (i.e., where nature chooses proposer) is

$$q_i^{h,t_i} = \frac{1}{m} \sum_{j \leq m} Q_i^{t_i}(h_j)$$

where m is the number of active agents, and h_j is the information set following h in which j is the proposer.

(4) We are now able to define the rationality constraints. We require that the payoff from the equilibrium behavioral strategy σ exceeds the payoffs of using pure strategies. Specifically, in PBE, for all i, $t_i \in T_i$, all h that correspond to one of i's information sets, and all actions $b \in A(h)$, we have:

$$\sum_{t_{-i}} \mu_i^h(t_{-i}) \sum_{a \in A(h)} \sigma_i^{h,t_i}(a) q_i^{h(t_{-i}),t_i}(a) \geq \sum_{t_{-i}} \mu_i^h(t_{-i}) q_i^{h(t_{-i}),t_i}(b)$$

We also add constraints for the Bayesian update of belief variables for any agent i regarding type t_j^κ of agent j performing a_j at any h (for all i, $t_i \in T_i$, all h and all a_j):

$$\mu_i^{h \cup a_j,t_i}(t_j^\kappa) = \mu_i^{h,t_i}(t_j^\kappa) \sigma_j^{h,t_j^\kappa}(a_j) / \sum_{t_j^k \in T_j} \mu_i^{h,t_i}(t_j^k) \sigma_j^{h,t_j^k}(a_j)$$

Finally, we add the obvious constraints specifying the domain of the various variables denoting strategies or beliefs (they take values in $[0, 1]$ and sum up to 1 as appropriate).

This ends the formulation of the program describing the PBE. This is a polynomial constraint satisfaction problem: finding a solution to this system of constraints is equivalent to deciding whether a system of polynomial equations and inequalities has a solution [1]. The problem is decidable, but is intractable. For example, an algorithm for deciding this problem has been proposed with exponential complexity [1]. Specifically, the complexity of deciding whether a system of s polynomials, each of degree at most d in k variables has a solution is $s^{k+1}d^{O(k)}$. In our case, assuming a random choice of proposer at each of S rounds, we can show that if α is the number of pure strategies, N the number of agents, T the number of types, then $s = O(N^S)$, $d = NS$ and $k = O(\alpha NT)$. This is due to a variety of combinatorial interactions evident in the constraints above, creating as they do interdependencies between belief and strategy variables.

In summary, the formulation above characterizes the PBE solution of our coalitional bargaining game as a solution of a polynomial program. However, it does not seem possible that this solution can be efficiently computed in general. Nevertheless, this PBE formulation may prove useful for the computation of a PBE in a bargaining setting with a limited number of agents, types, proposals and bargaining stages.

4 Approximations

The calculation of the PBE solution is extremely complex due to both the size of the strategy space (as a function of the size of the game tree, which grows exponentially with the problem horizon), and the dependence between variables representing strategies and beliefs, as explained above. We present an approximation strategy that circumvents these issues to some degree by: (a) performing only a small lookahead in the game tree in order to decide on a action at any stage of the game; and (b) fixing the beliefs of each agent during this process. This latter approach, in particular, allows us to solve the game tree by backward induction, essentially computing an equilibrium for this fixed-beliefs game. Note that while beliefs are held fixed during the lookahead (while computing an immediate action), they do get updated once the action is selected and executed, and thus do evolve based on the actions of others (this is in the spirit of receding horizon control). Furthermore, we allow sampling of type vectors in the computation to further reduce the tree size.

More precisely, at any stage of the game, with a particular collection of active agents (each with their own beliefs), we implement the following steps:

1. An agent (e.g., proposer) constructs a game tree consisting of the next d rounds of bargaining (for some small *lookahead* d).[2] All active agents are assumed to have fixed beliefs at each node in this tree corresponding to their beliefs at the current stage. The agent computes its optimal action for the current round using backward induction to approximate an equilibrium (similar in nature to an *SPE*) of this limited depth game. (We elaborate below.) Furthermore, they *sample* partners' types when calculating the values of coalitions and proposals.

[2] If less than d rounds remain, the tree is suitably truncated.

2. Each player executes its action computed for the current round of bargaining. If a coalition is formed, it breaks away, leaving the remaining players as active.
3. All active agents update their beliefs, given the observed actions of others in the current round, using Bayesian updating. Further, each agent keeps track of the belief updates that any other agent of a specific type would perform at this point.
4. The next bargaining round is implemented by repeating these steps until a complete coalition structure is determined or the maximum number of bargaining rounds is reached.

We stress that the algorithm above does not approximate the PBE solution; getting good bounds for a true PBE approximation would only be likely by assuming belief updating at *every* node of the game tree mentioned in Step 1. However, if our algorithmic assumptions are shared by all agents, each can determine their best responses to others' (approximately) optimal play, and thus their play approximates an equilibrium of the fixed-beliefs game. Indeed, we can define a *sequential equilibrium under fixed beliefs (SEFB)* as an extension of the SPE and a restriction of the PBE for a fixed-beliefs bargaining game, and can show the following (stated informally here):

Theorem 1 *If the Bayesian core (BC) of a Bayesian coalitional game G [2] is non-empty, and so is the BC of each one of G's subgames, then—regardless of nature's choice of proposers —there is an SEFB strategy profile of the corresponding fixed-beliefs discounted Bayesian coalitional bargaining game that produces a BC element; and conversely, if there is an order independent*[3] *SEFB profile for a Bayesian coalitional bargaining game, then it leads to a configuration that is in the BC of the underlying G.*

This result describes some notion of equivalence between cooperative and non-cooperative Bayesian coalition formation solution concepts, and is similar to results (e.g., Moldovanu et al. [5]) for non-stochastic environments. It also motivates further Step 1 of our heuristic algorithm, equating fixed belief equilibrium computation with determination of (i's part of) the Bayesian core. We now elaborate on this process.

We assume that the agents proceed to negotiations that will last d rounds (corresponding to the algorithm's lookahead value d) under the assumption that all beliefs will remain fixed to their present values throughout the (Step 1) process. We will present the deliberations of agent i during negotiations. For *fixed* types t_{-i} of possible partners, drawn according to μ_i, i will reason about the game tree and assume fixed beliefs of other agents. (Agents *will* track of the updates of other agents' beliefs after this stage of bargaining; see Step 2 above). Then, i can calculate the optimal action of any t_j agent (including himself) at any information set by taking expectations over the corresponding tree nodes.

We begin our analysis at the last stage d of negotiations. In any node ξ after history h where i of type t_i is a responder to proposal $\pi \in \mathcal{P}$ and assumes a specific type vector for partners, he expects a value for accepting that is different to his (discounted) reservation value only if all other responders accept the proposal as well:

$$q_i^{h(t_{-i}),t_i}(y) = \begin{cases} x_i V_d(t_C) & \text{if all } t_{-i} \in t_C \text{ accept} \\ V_d(t_i) & otherwise \end{cases} \quad (1)$$

[3] A strategy profile is *order independent* iff when played it leads to *a specific* $\langle CS, P\rangle$, independently of the choice of proposers.

However, to evaluate this acceptance condition, i would need to know the other responders' strategies (which in turn depend on i's strategy). Therefore, i will make the simplifying assumption that all other responders j evaluate their response to π by assuming that the rest of the agents (including i) will accept the proposal. Thus, any j with $t_j \in t_{-i}$ is assumed by i to accept if he evaluates his expected payoff from acceptance as being greater than his (discounted) reservation payoff:

$$x_j \sum_{t_{-j} \in t_C} \mu_j(t_{-j}) V_d(\{t_j, t_{-j}\}) \geq V_d(t_j) \quad (2)$$

With this assumption, i is able to evaluate the acceptance condition in Eq. 1 above, and so calculate a specific $q_i^{h(t_{-i}),t_i}(y)$ value. Note that the use of this assumption can sometimes lead to an overestimate of the value of a node.

At node $\xi = h(t_{-i})$, i can also evaluate his refusal value as $q_i^{h(t_{-i}),t_i}(n) = V_d(t_i)$ in this last round. Then, responder i's actual strategy at h can be evaluated as the strategy maximizing i's expected value given μ_i^{h,t_i}:

$$\sigma_i^{h,t_i} = \arg\max_{r \in \{y,n\}} \{ \sum_{t_{-i} \in t_C} \mu_i^{h,t_i}(t_{-i}) q_i^{h(t_{-i}),t_i}(r) \}$$

If i is a *proposer* of type t_i deliberating at $\xi = h(t_{-i})$, the value of making proposal π is:

$$q_i^{h(t_{-i}),t_i}(\pi) = \begin{cases} x_i V_d(t_C) \text{ if } \sigma_j^{h,t_j} = y, \ \forall j \\ V_d(t_i) \quad otherwise \end{cases} \quad (3)$$

(i.e., i will get his reservation value unless all the responders of the specific type configuration agree to this proposal). Furthermore, i's expected value $q_i^{h,t_i}(\pi)$ from making proposal π to coalition C at h can be determined given μ_i^{h,t_i}. Thus, the best proposal that i of type t_i can make to coalition C is the one with maximum expected payoff: $\sigma_i^{C;h,t_i} = \arg\max_\pi q_i^{h,t_i}(\pi)$ with expected payoff $q_i^{C;h,t_i}$.

However, i can also propose to other coalitions at h as well. Therefore, the coalition C^* to which i should propose is the one that guarantees him the maximum expected payoff: $C^* = \arg\max_C \{q_i^{C;h,t_i}\}$. If P^* is the payoff allocation associated with that proposal, then the optimal coalition-allocation pair that t_i can propose in this subgame (that starts with i proposing at h) is: $\sigma_i^{*;h,t_i} = \{C^*, P^*\}$ with maximum expected payoff $q_i^{C^*;h,t_i}$. Finally, if there exist more than one optimal proposal for i, i randomly selects any of them (this is taken into account in agents' deliberations accordingly).

Of course, when the subgame starts an agent i does not know who the proposer in this subgame will be; and i has only probabilistic beliefs about the types of his potential partners. Thus, i has to calculate his *continuation payoff* $q_i^{d:\xi,t_i}$ at stage d (that starts at node ξ) with m participants, in the way explained in the previous section. This is straightforward, as i can calculate his expected payoffs from participating in any subgame where some j proposes, given that any i can calculate the optimal strategies (and associated payoffs) for any j in this round d subgame.

Now consider play in a subgame starting in period $d-1$, again with the participation of m agents. The analysis for this round can be performed in a way completely similar to

the one performed for the last round of negotiations. However, there is one main difference: the payoffs in the case of a rejection are now the continuation payoffs (for agents of specific type) from the last round subgame. We have to incorporate this difference in our calculations. Other than that, we can employ a similar line of argument to the one used for identifying the equilibrium strategies in the last period. Proceeding in this way, we define the continuation payoffs and players' strategies for each prior round, and finally determine the first round actions for any proposer i of type t_i or any responder j of type t_j responding to any proposal.

5 Experimental Evaluation

To evaluate our approach, we first conducted experiments in two settings, each with 5 agents having 5 possible types. Agents repeatedly engage in *episodes* of coalition formation, each episode consisting of a number of negotiation rounds. We compare our Bayesian equilibrium approximation method (*BE*) with *KST*, an algorithm inspired by a method presented by Kraus et al. [4]. Though their method is better tailored to other settings, focusing on social welfare maximization, it is a rare example of a successfully tested discounted coalitional bargaining method under some restricted form of uncertainty, which combines heuristics with principled game theoretic techniques. It essentially calculates an approximation of a kernel-stable allocation for coalitions that form in each negotiation round with agents intentionally compromising part of their payoff in order to form coalitions. Like [4], our *KST* uses a compromise factor of 0.8, but we assume no central authority, only one agent proposing per round, and coalition values estimated given type uncertainty.

During an episode, agents progressively build a coalition structure and agree on a payment allocation. The action executed by a coalition at the end of an episode (the *coalitional action*) results in one of three possible stochastic outcomes $o \in O = \{0, 1, 2\}$ each of differing value. Each agent's type determines its "quality" and the "quality" of a coalition is dictated by the sum of the quality of its members less a penalty for coalition size.[4] Coalition quality then determines the odds of realizing a specific outcome (higher quality coalitions have greater potential). Finally, the value of a coalition given member types is the expected value w.r.t. the distribution over outcomes.

In our first setting, singleton coalitions receive a penalty of -1 quality points. We compare BE and KST under various learning models by measuring average total reward garnered by all coalitions in 30 runs of 500 formation episodes each, with a limit of 10 bargaining rounds per episode and a bargaining discount factor of $\delta = 0.9$. We also compare average reward to the reward that can be attained using the optimal, fixed "kernel-stable" coalition structure $\{\langle 1, 2, 3, 4\rangle, \langle 0\rangle\}$.

We compared BE and KST using agents that update their prior over partner types after observing *coalitional* actions—thus learning by reinforcement (*RL*) after each episode—and those that do not (*No RL*). In all cases, BE agents update their beliefs after observing the bargaining actions of others during each negotiation round. There are 388 proposals a BE agent considers when negotiating in a stage with all five agents present (fewer in other cases).

Table 1(a) shows performance when each agent has a uniform prior regarding the types of others. The *BE* algorithm consistently outperforms *KST*, even though *KST* promotes social welfare (i.e., is well-aligned with total reward criterion) rather than individual rationality. KST agents without RL always converge to the coalition structure $\{\langle 4\rangle, \langle 3\rangle, \langle 2\rangle, \langle 0, 1\rangle\}$; this is due to the fact that they are discouraged from cooperating due to the lack of information about their counterparts. When KST agents learn from observed actions after each episode (KST-Uni-RL) they form the coalitions $\{\langle 2, 3, 4\rangle, \langle 0\rangle, \langle 1\rangle\}$ in the last episode in 16 of 30 runs. BE agents, in contrast, form coalitions based on evolving beliefs about others, and do not form the optimal structure $\{\langle 1, 2, 3, 4\rangle, \langle 0\rangle\}$.[5] Rather they tend to form coalitions of 2 or 3 members which exclude agent 0 from being their partner. In addition, payoff division for BE agents is more aligned with individual rationality than it is with KST. The shares of (averaged) total payoff of KST-Uni-RL agents 0–4 are $0.8\%, 0.7\%, 28.8\%, 29.6\%, 40.1\%$, respectively, while for BE-Uni-RL (SS:10, LA:2) they are $1.3\%, 13.4\%, 18.8\%, 29.5\%, 37\%$; this more accurately reflects the *power* [6] of the agents. BE results are reasonably robust with changing sample size and lookahead value (at least in this environment with 3125 possible type vectors in a 5-agent coalition).

We attribute the poor performance of KST agents to the fact that they make their proposals without in any way taking into consideration the changing beliefs of others. With the beliefs of the agents varying, negotiations drag (up to the maximum of 10 rounds) due to refusals, resulting in reduced payoffs. BE agents do not suffer from this problem, since they keep track of all possible partners' updated beliefs, and use them during negotiation. Thus, they typically form a coalition structure within the first four rounds of an episode.

We also experimented with a second setting in which singleton coalitions receive a penalty of -2 quality points (rather than -1 above), and where $q(\vec{t}_C) = \sum_{t_i \in \vec{t}_C} q(t_i)/|C|$ (as coalitions get bigger they get penalized to reflect coordination difficulties). This setting makes the quality of coalitions more difficult to distinguish. Here, a near-optimal configuration contains the structure $\{\langle 4, 3\rangle, \langle 2, 1\rangle, \langle 0\rangle\}$. We use three different priors: *uniform*, *misinformed* (agents have an initial belief of 0.8 that an agent with type t has type $t + 2$), and *informed* (belief 0.8 in the true type of each other agent).

The results (Table 1(b)) indicate that KST agents again do not do very well, engaging in long negotiations due to unaccounted-for differences in beliefs among the various agents. KST-Uni-RL agents, for example, typically use all ten bargaining rounds; in contrast, BE-Uni-RL usually form structures within 3 rounds. Even when KST uses informed priors, the fact that the expected value of coalitions is not common knowledge takes its toll. BE agents, on the other hand, derive the true types of their partners with

[4] We omit the details here. We only note that agent 0 (of type 0) is detrimental to any coalition (in our 2 first settings).

[5] Nor should they, given bargaining horizon and δ—the kernel and other stability concepts do not consider bargaining dynamics.

Method	Reward
"Optimal" CS	65800 (expected)
KST-Uni-NoRL	32521.3(**49.4**%)
KST-Uni-RL	44274.4(**67.3**%)
BE-Uni-NoRL SS=20, LA=3	60037.7(**91.2**%)
BE-Uni-RL SS=20, LA=3	57775.4(**87.8**%)
BE-Uni-NoRL SS=10, LA=2	61444.3(**93.4**%)
BE–Uni-RL SS=10, LA=2	60086.7(**91.3**%)
BE-Uni-NoRL SS=3, LA=2	61269(**93.1**%)
BE-Uni-RL SS=3, LA=2	60301.1(**91.6**%)

(a) Setting A

Method	Reward
"Optimal" CS	33890 (expected)
KST-Uni-NoRL	20201.4(**59.6**)%
KST-Uni-RL	20157.7(**59.5**)%
BE-Uni-NoRL	31762.1(**93.7**%)
BE-Uni-RL	32275.9(**95.2**%)
KST-Mis-NoRL	20193.2(**59.6**)%
KST-Mis-RL	21642.5(**63.9**)%
BE-Mis-NoRL	31716.6(**93.5**%)
BE-Mis-RL	32293.7(**95.3**%)
KST-Inf-NoRL	22241.5(**65.6**%)
KST-Inf-RL	24748.1(**73**%)
BE-Inf-NoRL	31688.3(**93.3**%)
BE-Inf-RL	32401(**95.6**%)

(b) Setting B; (BE uses SS=10, LA=2)

Method	Q	A/B
KST-NoRL-0.95	2.15383	1.17
BE-NoRL-0.95	3.7698	1.71
KST-NoRL-0.5	6.88	4.26
BE-NoRL-0.5	8.4	1.5
KST-RL-0.95	2.20384	2.25
BE-RL-0.95	4.83322	1.7
KST-RL-0.5	2.96	3.15
BE-RL-0.5	9.96	1.43

(c) Setting C; Uniform Priors; BE uses SS=5, LA=2; A/B denotes *observed* relative power of A over B

Table 1: Settings' results (average)."SS":sample size;"LA":lookahead;"Uni":uniform,"Mis":misinformed,"Inf":informed prior.

certainty in all experiments, and typically form profitable configurations with structures such as $\{\langle 4,3\rangle, \langle 2,1\rangle, \langle 0\rangle\}$ or $\{\langle 4,2\rangle, \langle 3,1\rangle, \langle 0\rangle\}$. We can also see that RL enhances the performance of BE agents slightly, helping them further differentiate the quality of various partners.

We also report briefly on the results in a setting with 8 agents, of 2 possible types per agent (4 agents of type A, 4 of type B). The relative power of type A over B is 1.5.[6] In this setting, forming coalitions by mixing agent types is detrimental, with the exception of the $\langle A, A, B, B\rangle$ ("optimal"), $\langle A, A, B\rangle$ and $\langle A, B\rangle$ coalitions. There are 2841 proposals an agent considers when negotiating in a stage with all 8 agents present. The setting makes discovery of opponent types difficult, and thus rational agents should settle for suboptimal coalitions (hopefully using them as stepping stones to form better ones later). We also varied the bargaining δ (0.95 and 0.5). Agents do not accumulate much reward in this setting, bargaining for many rounds. Instead of reporting reward, we report *expected value Q of formation decisions*, $Q = \sum_C f_C V(C)$, with f_C being the observed average frequency with which coalition C forms and $V(C)$ its expected value. Results (Table 1(c)) show that BE agents outperform KST agents both in terms of social welfare and individual rationality (the observed relative power of types—the fraction of respective observed payoffs—is close to the true power), and that RL updates are quite beneficial. Further, lowering the discount rate to 0.5 forces the agents to form coalitions early, but also contributes to better decisions, because it enables the agents to discover the types of opponents with more accuracy, effectively reducing the number of possible opponent responses during bargaining (intuitively, given more time, both a "strong" and a "weak" type might refuse a proposal, while if time is pressing the "weak" might be the *only* one to accept).

6 Concluding Remarks and Future Work

We proposed an algorithm for coalitional bargaining under uncertainty about the capabilities of potential partners. It uses iterative coalition formation with belief updating based on the observed actions of others during bargaining, and is motivated by our formulation of the PBE solution of a coalitional bargaining game. The algorithm performs well empirically, and can be combined with belief updates after observing the results of coalitional actions (in reinforcement learning style).

Future and current work includes implementing a continuous bargaining action space version of our algorithm, and also incorporating it within a broader RL framework facilitating coalition formation and sequential coalitional decision making under uncertainty. We are also investigating approximation bounds for our heuristic algorithm.

[6]Relative power A/B is the expected payoff of A in coalitions excluding B, over the expected payoff of B in coalitions without A.

Acknowledgments

Thanks to Vangelis Markakis for extremely useful discussions and helpful comments.

References

[1] S. Basu, R. Pollack, and M.-F. Roy. On the Combinatorial and Algebraic Complexity of Quantifier Elimination. *Journal of the ACM*, 43(6):1002–1045, 1996.

[2] G. Chalkiadakis and C. Boutilier. Bayesian Reinforcement Learning for Coalition Formation Under Uncertainty. In *Proc. of AAMAS'04*, 2004.

[3] K. Chatterjee, B. Dutta, and K. Sengupta. A Noncooperative Theory of Coalitional Bargaining. *Review of Economic Studies*, 60:463–477, 1993.

[4] S. Kraus, O. Shehory, and G. Taase. The Advantages of Compromising in Coalition Formation with Incomplete Information. In *Proc. of AAMAS'04*, 2004.

[5] B. Moldovanu and E. Winter. Order Independent Equilibria. *Games and Economic Behavior*, 9, 1995.

[6] R.B. Myerson. *Game Theory: Analysis of Conflict*. 1991.

[7] A. Okada. A Noncooperative Coalitional Bargaining Game With Random Proposers. *Games and Econ. Behavior*, 16, 1996.

[8] M.J. Osborne and A. Rubinstein. *A course in game theory*. 1994.

[9] J. Suijs, P. Borm, A. De Wagenaere, and S. Tijs. Cooperative games with stochastic payoffs. *European Journal of Operational Research*, 113:193–205, 1999.