Web Intelligence and Agent Systems: An International Journal 0 (2012) 1–0
IOS Press

Sequential Multi-Agent Exploration for a Common Goal

Igor Rochlin [a], David Sarne [a] and Gil Zussman [b]
[a] *Department of Computer Science*
Bar-Ilan University
Ramat-Gan, Israel
E-mail: {igor.rochlin,david.sarne}@gmail.com
[b] *Department of Electrical Engineering*
Columbia University
New York, NY
E-mail: gil@ee.columbia.edu

Abstract. Motivated by applications in Dynamic Spectrum Access Networks, we focus on a system in which a few agents are engaged in a costly individual exploration process where each agent's benefit is determined according to the minimum obtained value. Such an exploration pattern is applicable to many systems, including shipment and travel planning. This paper formally introduces and analyzes a sequential variant of the general model. According to that variant, only a single agent engages in exploration at any given time, and when an agent initiates its exploration, it has complete information about the minimum value obtained by the other agents so far. We prove that the exploration strategy of each agent, according to the equilibrium of the resulting Stackelberg game, is reservation-value based, and show how the reservation values can be calculated. We also analyze the agents' expected-benefit maximizing exploration strategies when they are fully cooperative (i.e., when they aim to maximize the expected joint benefit). The equilibrium strategies and the expected benefit of each agent are illustrated using a synthetic homogeneous environment, thereby demonstrating the properties of this new exploration scheme and the benefits of cooperation.

Keywords: Multi-Agent Exploration, Multilateral Search, Cooperation, Dynamic spectrum access networks, Game Theory

1. Introduction

This paper focuses on exploration problems stemming from the spectrum sensing process of users in a Cognitive Radio Network (also known as Dynamic Spectrum Access Network). A Cognitive Radio was first defined by Mitola [35] as a radio that can adapt its transmitter parameters to the environment in which it operates. According to the Federal Communications Commission (FCC), a large portion of the assigned spectrum is used only sporadically [11]. Due to their adaptability and capability to utilize the wireless spectrum opportunistically, Cognitive Radios are considered key enablers for efficient use of the spectrum [1,12,13].

Under the basic model of Dynamic Spectrum Access Networks [1], Secondary Users (SUs) can use *white spaces* that are not used by the Primary Users (PUs) but must avoid interfering with active PUs.[1] In order to identify available PU channels, the SUs have to sense the spectrum and to obtain the quality of the different available channels. In particular, a *spectrum sensing* mechanism has to determine how and when the SUs sense the different channels, and a *spectrum decision* mechanism has to determine which channel best satisfies the application requirements (different channels may have different qualities) [1,17]. While spectrum sensing is primarily a physical layer

[1] PUs and SUs are also referred to as Licensed and Opportunistic Users, respectively.

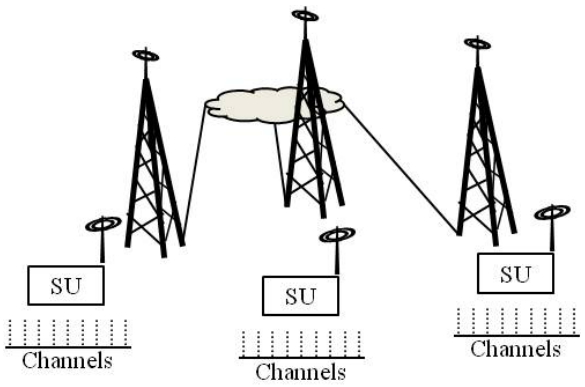

Fig. 1. An illustration of the exploration process of the SUs which are connected via an infrastructure network.

issue, we focus on the functionalities above the physical layer that determine how the sensing should be performed.

Specifically, we focus on the exploration process of SUs which is illustrated in Figure 1.[2] In our model, there are a few non-interfering SUs that are served by different base stations or access points. These SUs need to maintain a connection through the infrastructure network. Each of the SUs has to sense the channels in its own environment (i.e., to its own base station) and to select a specific channel based on the channels' qualities (these qualities are not known in advance as they depend on PUs' activity). Sensing a specific channel consumes the SU's resources (e.g., energy), and therefore an SU will usually not sense all the channels in its environment.

The key point is that since all SUs participate in the connection, the overall performance of the system (e.g., bandwidth allocated to the connection) is a function of the quality of the worst channel. Each SU's decision of whether or not to sense and obtain an additional channel should thus be based on the tradeoff between the expected incremental improvement that can be obtained in the connection's quality and the cost (in terms of energy spending) associated with the additional sensing operation. The fact that the marginal improvement also depends on the findings of the other agents substantially complicates the calculation of the SUs' exploration strategy. Particularly, when the SUs are self-interested, the set of exploration strategies should be derived based on equilibrium analysis.

[2]We note that modeling the effects of the PU arrival process [23] is out of the scope of this work.

Models of agents engaged in costly exploration processes involving the evaluation of different available options (opportunities), out of which only one needs to be chosen, are quite common in multi-agent systems (MAS) [7,26,27,19]. In these models, the goal of the agent is not necessarily to find the opportunity associated with the maximum value but rather to maximize the overall benefit, defined as the value of an opportunity eventually picked minus the costs accumulated during the exploration process. Economic search theory provides a framework for optimizing the performance of an agent in such costly exploration settings [32,40,14]. The expected-benefit maximizing exploration strategy in such models is commonly reservation-value based. Namely, the agent obtains opportunities as long as the best value found is lesser (or greater, depending on the application) than a pre-defined threshold.

Despite the richness of research of costly exploration, the models used commonly assume a *single* exploration process. Yet, as illustrated by the Dynamic Spectrum Access example, in reality agents often need to take into consideration the performance of other agents engaged in the exploration process. A similar example for such a dependency exists in "Mars rovers"-like applications. For example, when the robots need to evaluate different routes in order to get to a pre-defined location in order to mine a certain mineral on the face of Mars. The evaluation of different routes is costly as the robots possibly need to gather supplementary information required for the process. If all agents need to be present at the location in order for the mining process to be executed, then their performance depends on the maximum among the minimum individual times it takes any of them to arrive. Another such typical scenario can be found in coordination-management applications (e.g., DARPA's Coordinators project) where the quality of performance of a task is commonly defined by a quality accumulation function over the sub-tasks [48]. This also holds in complex transportation problems, e.g., those that involve ground, air and sea transportation. Assume that for each segment of the shipment route different offers should be received from shipment companies. The selection of a container/vehicle for each segment of the route dictates the amount and type of cargo that can be transported overall (a single bottleneck). Similarly, when planning a trip and requesting quotes from different vendors (e.g., for flights and accommodation), the correlation between the results of different explo-

ration efforts substantially influences the overall performance.

In this paper, we formally introduce the model of a multi-agent exploration in which different agents need to explore individually and the performance of each agent is affected by the performance of all the others. In particular, we focus on a protocol under which only one agent explores at a time, and each agent's exploration starts upon receiving the final outcome of all agents that have explored prior to it. Since the agents are exploring sequentially, the problem can be considered from a game theory perspective and formulated as a Stackelberg game, where each agent is a follower for the agent exploring prior to it and a leader for the agent performing the subsequent exploration process. We provide a comprehensive analysis of the problem and prove that the equilibrium strategies for the sequential multi-agent exploration are reservation-value based. Based on the analysis, we obtain the appropriate equations from which the reservation value of each opportunity can be calculated. Complementary analysis for the case of fully cooperative agents and defection from cooperation scenarios is also provided. We use homogeneous environments (i.e., where all opportunities available to an agent share the same exploration cost and probabilistic properties) to illustrate the effect of different parameters on the equilibrium exploration strategies and the expected benefit of the different agents. Preliminary results of the research reported in this paper appear in [39].

2. The Model

We consider a setting where k individual agents need to establish an ad-hoc partnership from which they all benefit.[3] Each agent A_i sees a different value in the partnership, denoted v_i. The value v_i, seen by agent A_i in the partnership, is the result of an exploration process involving n_i opportunities, denoted $O_i = \{o_i^1, .., o_i^{n_i}\}$, from which the agent needs to choose one. While the value of each opportunity o_i^j is a priori unknown to agent A_i, the agent is acquainted with the probability distribution function $f_i^j(y)$ from which it is derived. In order to obtain the true value of opportunity $o_i^j \in O_i$, agent A_i needs to consume some of its resources. This is modeled by the cost c_i^j, expressed in terms of opportunity values. Therefore, the agents are required to conduct an exploration process which takes into consideration the tradeoff between the marginal improvement in the value they see in the partnership and the accumulated cost incurred along the process.

The value v_i that agent A_i sees in the partnership is therefore the maximum among the values obtained along its individual exploration process. This value, however, is only an upper bound to the value that agent A_i can potentially gain from the partnership. The actual value each of the agents gains from the partnership, denoted v^* (and termed "effective value" hereafter), is the minimum of the values seen by all agents in the partnership, i.e., $v^* = \min\{v_i | i = 1, .., k\}$. The model assumes full information in the sense that all agents are acquainted with all of the distribution functions $f_i^j(y)$ and exploration costs c_i^j.

Taking the cognitive radio application domain as an example, each agent represents an SU and all SUs are interested in establishing a conference call, use a document/video sharing application or play in a multiplayer game. The SUs are located in different geographical locations and each SU can use different wireless channels to connect to a server supporting the requested application. Each SU senses several different channels of different qualities until it selects a specific channel with a specific quality (the SU spends some resources, e.g., energy, to sense each channel). The quality of service provided by application depends on the qualities of all individual channels (e.g., if one of the SUs has a low quality channel, the experience of all of the users will be negatively affected). Hence, the quality of service provided to all the SUs will be a function of the lowest quality channel selected by one of the SUs (See Figure 1).

Table 1 provides mappings of the cognitive radio cooperative exploration application and the different applications discussed in the former section to the model formally presented above. In all these examples the exploration itself consumes some resources, resulting in a tradeoff between the benefit from the potential improvement in the quality of the results that may be further obtained and the costs of the additional exploration. Also, in all these examples, the value from which the agents benefit is the minimum among the maximum values found in the individual explorations (except for the Mars rovers application in which the maximum among the minimum distances found individually is used, which is essentially a dual problem and thus equivalent).

[3] See Appendix B for a summary of all the notations used in this paper.

Table 1
Mapping real-life applications to the sequential exploration problem.

Application	Agent	Opportunity			Benefit
		Essence	Value	Exploration cost	
Cognitive radio	SUs	Channels to communicate with server	Quality of channel	Battery power required to sense additional channels	The lowest quality among the qualities of channels selected by the SUs
Mars rovers	Robots	Routes to destination point	Time to get to destination	Battery power required for communication and obtaining complementary data	Maximum among the minimum individual times it takes any of the robots to arrive to the location
Coordinators	Scheduling assistants	Alternative plans for task execution	Quality of plan	Resources required for schedule evaluation	Minimum quality achieved (when using minimum quality accumulated function)

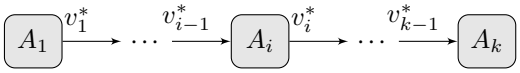

Fig. 2. The sequential k-agent exploration process. The figure illustrates schematically the sequential exploration process, where the exploration of agent A_i starts only after agent A_{i-1} has finished its exploration (where A_1 is the first to explore). When agent A_{i-1} finishes its exploration, it broadcasts the value $v_{i-1}^* = \min\{v_i | i = 1, .., i-1\}$ to the next agent in the sequence. The effective value is $v^* = \min\{v_i | i = 1, .., k\}$.

There are many protocols the agents can follow for executing this multi-agent exploration, differing in the parallelism and levels of cooperation along the process. For example, the agents can explore opportunities simultaneously, in parallel, with no interaction between them until each terminates its exploration process and shares the best value obtained. Another option is to take turns in exploration and share the values found along the process. Each exploration model variant is associated with different advantages, disadvantages and computational complexities. In this paper, we investigate a sequential exploration protocol by which agent A_i performs its exploration process as a whole only after agent A_{i-1} has finished its exploration (where A_1 is the first to explore) and the values obtained by agents finishing their exploration which become common knowledge (See Figure 2). The sequential nature of the process suggests that the exploration of the multiple agents is not iterative hence there are no convergence issues involved.

We distinguish between two principal settings, varying in the way each agent's expected benefit is defined. In the first, all agents are self-interested. Namely, each agent A_i attempts to maximize its own overall expected benefit, denoted EB_i, defined as the effective value, v^*, minus the expected accumulated cost incurred by that agent along its exploration. The problem in this case can therefore be thought of as a Stackelberg game, where each agent A_i is the first mover and wants to maximize its expected benefit with respect to its extent of exploration, which affects the strategies used by the remaining $k - i$ agents. In the second setting, all agents are cooperative, thus each agent's goal is to maximize the aggregate of all agents' expected benefits, i.e., kv^* minus the sum of the costs accumulated along all agents' explorations. The cooperative setting is common when the agents involved represent individuals from the same organization or family.

3. Analysis

We first introduce the expected-benefit maximizing exploration strategy for a single agent when the value it sees in the different opportunities is not constrained by the exploration of other agents. We then augment that strategy and adapt it to the case of k-agents with the minimum value restriction, distinguishing between the cooperative and the non-cooperative cases. The new strategies are proven to be expected-benefit maximizing for each agent, given the values it receives, resulting in the equilibrium set of strategies for the Stackelberg game.

The unique characteristics of the agents' expected-benefit maximizing exploration strategies and the resulting equilibrium dynamics in the self-interested variant of the multi-agent model are demonstrated using a simplistic setting where opportunities are homogeneous and the agents can explore as many opportunities as they request. The use of the homogeneous setting is more tractable numerically. It thus facilitates the illustration of the main differences between the sequential multi-agent exploration strategies model and others the agents may use, as well as the differ-

ences between the strategies of the different agents under different cooperation schemes. Specifically, we use a setting where all opportunities available to agent A_i $(i = 1, ..k)$ share the same exploration cost and probability distribution function, denoted c_i and $f_i(y)$ respectively. For the latter we use a uniform distribution function defined over the interval (a, b) (i.e., $f_i(y) = 1/(b-a), \forall a \leq y \leq b$, otherwise $f_i(y) = 0$). Any setting of this type can therefore be represented as $\{(c_1, (a_1, b_1)), .., (c_k, (a_k, b_k))\}$. We stress that even though such a setting is standard in costly exploration literature [32,30], its use in our case is merely for illustration purposes and all the results given in this paper are based on formal theoretical proofs.

3.1. Optimal Exploration with no Restrictions on the Values

When relaxing the restriction over the value obtained by the agent, each agent's exploration process can be analyzed separately and solved as an optimization problem. The individual exploration problem in this case can be mapped to the canonical exploration problem described by Weitzman [52]. Weitzman's model considers a single agent facing a setting similar to the one used for each of the agents in our model, except that the agent's expected benefit is the highest value it finds minus the expected cost incurred along its exploration process.

The optimal (expected-benefit maximizing) exploration strategy in Weitzman's model is inherently sequential (i.e., exploring one opportunity at a time). It is based on setting a reservation value (a threshold), denoted r^j for each opportunity o^j. The reservation value r^j to be used should satisfy:

$$c^j = \int_{y=r^j}^{\infty} (y - r^j) f^j(y) dy \quad (1)$$

Intuitively, r^j is the value where the agent is precisely indifferent: the expected marginal benefit from exploring an opportunity (i.e., obtaining its value) exactly equals the cost of the exploration. The agent should always choose to obtain the value of the opportunity associated with the maximum reservation value and terminate the exploration once the maximum value obtained so far is greater than the maximum reservation value of any of the opportunities which has not yet been explored.

We denote the above exploration strategy as "naive" in the context of the multi-agent exploration with value restriction, since it does not take into consideration the exploration of the other agents. In the following paragraphs we investigate the expected-benefit maximizing exploration strategy of an agent given the exploration strategy of the other agents and the minimum value constraint. We show that this latter strategy is qualitatively similar to the "naive" one, i.e, carried out sequentially according to reservation values, though the reservation values used are different and the new strategy does not necessarily follow the same exploration sequence as in the "naive" case.

3.2. Expected-Benefit Maximizing Strategy for the Sequential Multi-Agent Exploration with Value Restriction

Our analysis of the multi-agent exploration distinguishes between the expected-benefit maximizing strategy of agent A_k (the last agent in the sequence) and those of the other agents. Unlike the other agents, agent A_k's exploration does not depend on any future explorations of other agents and its only input is the minimum value obtained by the former $k-1$ agents, denoted $v^*_{k-1} = \min\{v_i | i = 1, .., k-1\}$. Any other agent A_i $(1 \leq i < k)$ takes into account the strategy of the remaining $k-i$ agents in its exploration strategy, in addition to the minimum value obtained by the former $i-1$ agents (denoted $v^*_{i-1} = \min\{v_i | i = 1, .., i-1\}$, where $v^*_0 = \infty$).

While many domains dictate a sequential exploration in the agent level (e.g., spectrum sensing technology precludes the evaluation of several channels simultaneously), the superiority of this approach over parallel (or partially-parallel) exploration is quite straightforward: exploring the opportunities in a subset $\bar{O}_i \subset O_i$ in parallel is equivalent to a sequential exploration where the decision is made at the end of the process. Therefore, the optimal sequential exploration strategy results in an expected benefit at least as good as the parallel one.

In the following paragraphs we prove that the individual expected-benefit maximizing strategy of each agent (best-response), given the input it receives (in terms of prior agents' findings), is reservation-value based. Consequently each strategy which is part of the equilibrium set of strategies of the resulting Stackelberg game is also of this structure.

We begin by developing the expected-benefit maximizing exploration strategy for agent A_k, given the value v^*_{k-1} received based on the exploration of the other $k-1$ agents. Obviously, if A_k obtains a value

greater than v^*_{k-1} along its own exploration then it necessarily terminates its exploration, as its benefit from the partnership cannot be improved further.

We use $(v^*_{k-1}, w, \bar{O}_k)$ to denote the state of agent A_k, where v^*_{k-1} is the minimum value obtained by the previous $k-1$ agents that have already finished their exploring, w is the best value found so far by A_k and $\bar{O}_k$ ($\bar{O}_k \subseteq O_k$) is the set of opportunities which values have not yet been obtained.

We denote the expected value the agents end up with at the end of the process as a whole (the "effective value") given that agent A_k is currently in state $(v^*_{k-1}, w, \bar{O}_k)$ by $E_k[v^*|(v^*_{k-1}, w, \bar{O}_k)]$. The expected effective value, if agent A_k is about to start its exploration process after receiving a value v^*_{k-1}, denoted $E_k[v^*|v^*_{k-1}]$, is thus given by $E_k[v*|(v, -\infty, O_k)]$.

Theorem 1. *The expected-benefit maximizing exploration strategy for agent A_k, given its state $(v^*_{k-1}, w, \bar{O}_k)$, is to set a reservation value $r^j_k < v^*_{k-1}$ for each opportunity $o^j_k \in \bar{O}_k$, where r^j_k is derived from:*

$$c^j_k = \int_{y=r^j_k}^{\infty} (min(y, v^*_{k-1}) - r^j_k) f^j_k(y) dy \qquad (2)$$

Agent A_k should always choose to explore the opportunity $o^j_k \in \bar{O}_k$ associated with the maximum reservation value and terminate the exploration once the maximum value obtained so far, w, is greater than the maximum reservation value of any of the remaining unexplored opportunities.

Proof. See appendix A.1

One important implication of Theorem 1 is that the reservation value calculation does not depend on the value found, w, nor on the set of unexplored opportunities $\bar{O}_k$. This means that the agent only needs to calculate n_k reservation values and this can be done before the exploration process even begins.

Figure 3 illustrates the expected benefit of agent A_k as a function of the reservation value it uses for a setting $\{*, (c_k = 0.1, (0,1))\}$ when receiving a value $v^*_{k-1} = 0.8$ (agent A_k's strategy depends solely on the value v^*_{k-1} and the characteristics of the opportunities available to A_k). The Figure also depicts the c_k and $\int_{y=r_k}^{\infty} (\min(y, v^*_{k-1}) - r_k) f_k(y) dy$ curves, demonstrating that the two intersect in the expected-benefit maximizing strategy's reservation value, according to (2).

Based on Equation 2 we observe two important and somehow intuitive properties of the expected-benefit

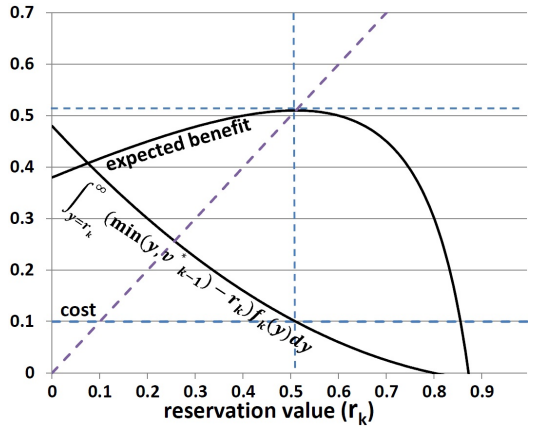

Fig. 3. The expected benefit of A_k as a function of the reservation value r_k it uses. The two other curves depict both sides of Equation 2, illustrating the intersection at the expected-benefit maximizing strategy's reservation value. The setting used is $\{*, (c_k = 0.1, (0,1))\}$.

maximizing strategy's reservation value r^j_k. First, r^j_k increases as v^*_{k-1} increases — the increase in v^*_{k-1} translates to an increase in $\min(y, v^*_{k-1})$, thus in order for Equation 2 to hold, r^j_k must also increase. The intuitive interpretation of this is that an increase in v^*_{k-1}, which is in fact an upper bound for the value each of the remaining agents may obtain from the partnership, should encourage each of the agents to increase its individual reservation value as greater values found could be exploited to a greater extent. The second observation is that the reservation value r^j_k decreases as c^j_k increases — the only way the right hand side of Equation 2 can increase is through a decrease in r^j_k. Intuitively, this can be explained by the fact that the increase in the cost of exploration results in a decrease in the benefit from future explorations, hence the reservation value (which directly affects the extent of exploration) decreases. Furthermore, Proposition 1 presents an important property of the correlation between r^j_k and v^*_{k-1}.

Proposition 1. *The difference between v^*_{k-1} and r^j_k increases as v^*_{k-1} increases, i.e., $0 < \frac{dr^j_k}{dv^*_{k-1}} < 1$.*

Proof. See appendix A.2.

Intuitively, Proposition 1 can be explained by the fact that when v^*_{k-1} is relatively small, agent A_k is likely to achieve an improvement similar to the increase in v^*_{k-1} within a small number of explorations. When v^*_{k-1} is relatively high, in order to fully match

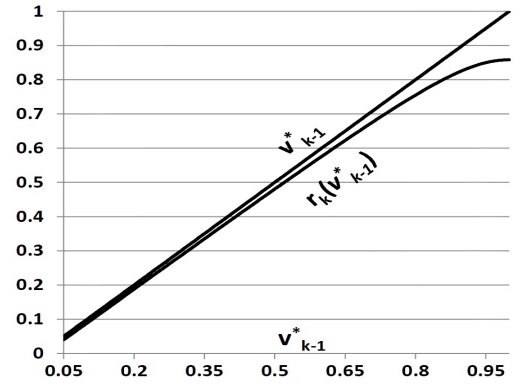

Fig. 4. The expected-benefit maximizing strategy's reservation value of A_k (according to (2)), as a function of v^*_{k-1}. The value of r_k increases in a decreasing rate as v^*_{k-1} increases. The setting used is $\{*, (c_k = 0.01, (0, 1))\}$.

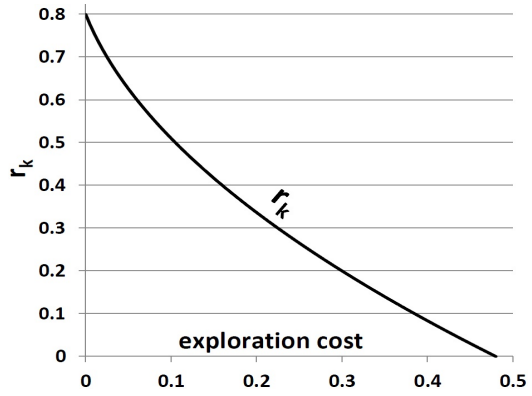

Fig. 5. The expected-benefit maximizing strategy's reservation value as a function of the exploration cost, c_k. The setting used is $\{*, (c_k, (0, 1))\}$ and $v^*_{k-1} = 0.8$.

the increase with an increase in the value the agent sees in the partnership, it needs to go through substantial exploration and consequently incurs substantial cost, which makes any increase in r^j_k less favorable.

Figure 4 depicts the correlation between r_k and v^*_{k-1} for a setting $\{*, (c_k = 0.01, (0, 1))\}$. The curve of $r_k(v^*_{k-1})$ complies with Proposition 1, i.e., increases as v^*_{k-1} increases, in a decreasing rate. Figure 5 depicts the correlation between r_k (calculated according to (2)) and c_k for the setting $\{*, (c_k, (0, 1))\}$ where $v^*_{k-1} = 0.8$. As expected, the value of r_k is always smaller than v^*_{k-1} and decreases as c_k increases.

We now move to analyze the expected-benefit maximizing strategy of the remaining agents. Similar to the case of the k-th agent, we use $(v^*_{i-1}, w, \bar{O}_i)$ to denote the state of agent A_i $(1 \leq i < k)$, where v^*_{i-1} is the minimum value obtained by the previous $i-1$ agents that have already finished their exploring, w is the best value found so far by A_i and $\bar{O}_i$ ($\bar{O}_i \subseteq O_i$) is the set of opportunities which values have not yet been obtained. The expected effective value, if agent A_i is about to start its exploration process after receiving a value v^*_{i-1}, denoted $E_i[v^*|v^*_{i-1}]$, is thus given by $E_i[v^*|v] = E_i[v^*|(v, -\infty, O_i)]$. Using the new notations, we can now introduce Theorem 2, which specifies the expected-benefit maximizing strategy for any agent A_i $(1 \leq i < k)$ given its state.

Theorem 2. *The expected-benefit maximizing exploration strategy for each agent A_i $(i = 1, .., k-1)$ when in state $(v^*_{i-1}, w, \bar{O}_i)$ is to set a reservation value r^j_i for each opportunities $o^j_i \in \bar{O}_i$, where r^j_i derives from:*

$$c^j_i = \int_{y=r^j_i}^{\infty} (E_{i+1}[v^*|min(v^*_{i-1}, y)] - E_{i+1}[v^*|r^j_i])f^j_i(y)dy \quad (3)$$

where:
$E_{i+1}[v^*|v^*_i] = E_{i+1}[v^*|(v^*_i, -\infty, O_{i+1})]$,

$$E_i[v^*|(v^*_{i-1}, w, \bar{O}_i)] = \int_{y=-\infty}^{r^j_i} E_i[v^*|(v^*_{i-1}, max(w, y), \bar{O}_i - o^j_i)]f^j_i(y)dy + \int_{y=r^j_i}^{\infty} E_{i+1}[v^*|min(v^*_{i-1}, max(w, y))]f^j_i(y)dy, \quad (4)$$

$$E_i[v^*|(v^*_{i-1}, w, null)] = E_{i+1}[v^*|min(v^*_{i-1}, w)],$$

$E_1[v^*|v^*_0] = E_1[v^*|\infty]$
and
$E_{k+1}[v^*|v^*_k] = v^*_k$
The agent should choose to explore the opportunity $o^j_i \in \bar{O}_i$ associated with the maximum reservation value and terminate the exploration process once the maximum value, w, obtained so far is greater than the maximum reservation value of any of the remaining unexplored opportunities.

Proof. See appendix A.3

Similar to the case of A_k, the reservation value calculation of agent A_i $(1 \leq i < k)$ does not depend on the value found, w, nor on the set of unexplored opportunities $\bar{O}_i$. This means that the agent only needs to calculate n_i reservation values and this can be done even before the exploration process even begins. Fur-

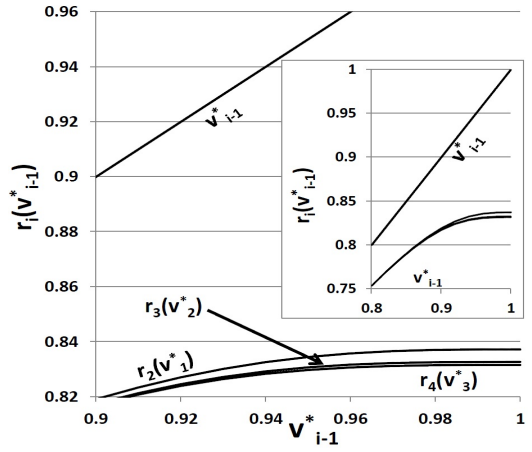

Fig. 6. The expected-benefit maximizing strategy's reservation value of A_i $(1 < i < 5)$ (according to (3)), as a function of v^*_{i-1}. The value of r_i increases in a decreasing rate as v^*_{i-1} increases. The setting used is $\{*,(c_2 = 0.01,(0,1)),(c_3 = 0.01,(0,1)),(c_4 = 0.01,(0,1)),(c_5 = 0.01,(0,1))\}$. The small figure to the right depicts the same curves over a wider range of v^*_{i-1} values.

thermore, based on Equation 3 we can draw the same conclusions regarding the correlation between r^j_i, v^*_{i-1} and c^j_i for any agent A_i $(1 < i < k)$ as given for the case of agent A_k. This time, however, the proof of correctness relies on the fact that $E_{i+1}[v^*|\min(v^*_{i-1}, y)]$ increases as v^*_{i-1} increases (this is proven as part of Theorem 2's proof). Given the latter property of $E_{i+1}[v^*|\min(v^*_{i-1}, y)]$, an increase in v^*_{i-1} requires an increase in r^j_i in order for Equation 3 to hold. Similarly, an increase in c^j_i requires a decrease in r^j_i.

Figure 6 depicts the correlation between r_i $(1 < i < 5)$ and v^*_{i-1} for a setting $\{*, (c_2 = 0.01, (0,1)), (c_3 = 0.01, (0,1)), (c_4 = 0.01, (0,1)), (c_5 = 0.01, (0,1))\}$. The curve $r_i(v^*_{i-1})$ complies with the above properties, i.e., increases as v^*_{i-1} increases at a decreasing rate. This is explained by the fact that the reservation values that agent A_i assigns to its opportunities are always less than the value v^*_{i-1} it receives as an input for its exploration.

The proof of the $0 < \frac{dr^j_i}{dv^*_{i-1}} < 1$ property for the general case (i.e., for any $1 < i < k$) uses the same principles used for proving Proposition 1, taking advantage of the relation $E_i[v^*|v^*_{i-1}] < E_i[v^*|v^{*'}_{i-1}]$ for any two values $v^*_{i-1} < v^{*'}_{i-1}$.

The expected effective value obtained eventually by each agent from the partnership, $E[v^*]$, can be calculated using a recursive equation similar to (4):

$$E[v^*] = E_1[v^*|\infty] \tag{5}$$

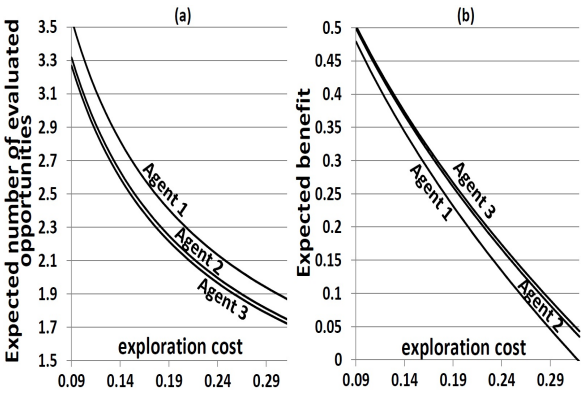

Fig. 7. (a) The expected number of evaluated opportunities; (b) Individual expected benefit of each agent as a function of the exploration cost c. The setting used is $\{(c_1 = c,(0,1)),..,(c_3 = c,(0,1))\}$.

Since the effective value v^* applies to all agents, the expected benefit of each agent differs only in its accumulated cost component. The expected cost of agent A_i, given the value it receives v^*_{i-1}, denoted $EC_i[cost|v^*_{i-1}]$, can be calculated using $EC_i[cost|v^*_{i-1}] = \sum_{j=1}^{n_i} c^j_i \prod_{l=1}^{j-1} F^l_i(r^j_i)$ (where F^l_i is cumulative distribution function of opportunity o^j_i). The term $\prod_{l=1}^{j-1} F^l_i(r^j_i)$ denotes the probability that A_i will eventually obtain, along its exploration process, the value of the opportunity associated with the j^{th} highest reservation value. The calculation of the a priori expected accumulated cost (that takes into consideration all possible v^*_{i-1} values), denoted $EC_i[cost]$, should weight $EC_i[cost|v^*_{i-1}]$ according to the probability of receiving each v^*_{i-1} value. This latter probability can be calculated using the same principles used in Equation 4. The expected benefit of any agent A_i, denoted EB_i, is thus given by:

$$EB_i = E[v^*] - EC_i[cost] \tag{6}$$

Figure 7 depicts the expected number of opportunities evaluated by the different agents (7(a)) and the expected benefit of the different agents (7(b)) in the setting $\{(c_1 = c,(0,1)),..,(c_3 = c,(0,1))\}$. As expected, the expected benefit decreases as the exploration cost increases. An interesting property of this symmetric setting is that the expected number of opportunities evaluated by A_i is less than those evaluated by A_{i-1}. This is explained by the fact that when all parameters are alike, $r_i < r_{i-1}$ and the agent associated with the lower reservation value terminate its exploration earlier than one using a greater reservation value. Similarly, in our symmetric case the ex-

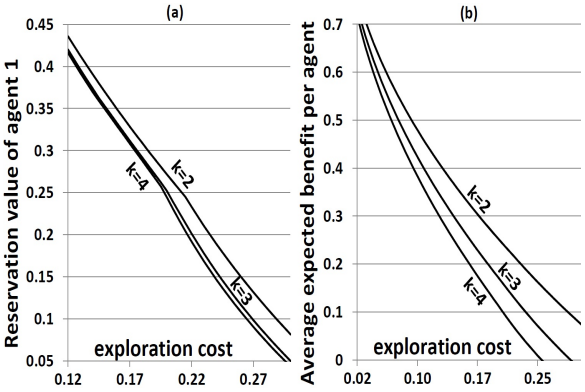

Fig. 8. (a) The expected-benefit maximizing strategy's reservation value of agent A_1 as a function of the exploration cost for different amounts of agents. (b) The average expected benefit (per agent) as a function of the exploration cost c for different amounts of agents. The setting used is $\{(c_1 = c, (0, 1)), .., (c_4 = c, (0, 1))\}$.

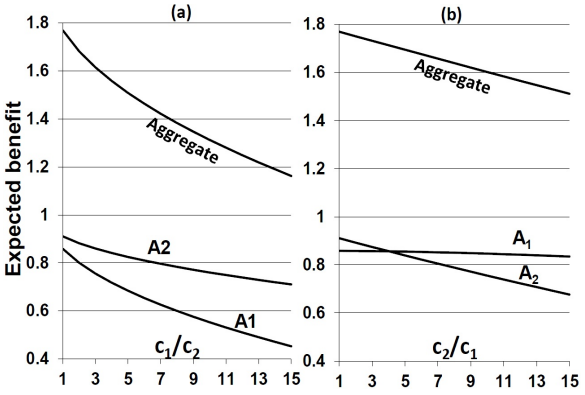

Fig. 9. The expected benefit of the agents when using sequential multi-agent exploration as a function of the exploration costs ratios: (a) c_1/c_2 (taking $c_2 = 0.01$, fixed); and (b) c_2/c_1 (taking $c_1 = 0.01$, fixed). The setting used is $\{(c_1, (0, 1)), (c_2, (0, 1))\}$.

pected benefit of agent A_i is greater than the expected benefit of A_{i-1} as they both enjoy the same effective value v^*, however, A_i "spends" less on exploration. We emphasize that these two latter properties (EB_i and expected number of opportunities evaluated by A_i decrease in i) do not necessary hold in general. For example, when using the setting $\{(c_1 = 0.02, (0, 10)), (c_2 = 0.02, (0, 1))\}$, the expected benefits are 0.88 and 0.8 and the expected number of evaluated opportunities are 1.1 and 5 for agents A_1 and A_2, respectively.

Figure 8 illustrates the effect of the number of agents involved in the exploration process on the reservation values used and the average expected benefit per agent when using the setting $\{(c_1 = c, (0, 1)), .., (c_4 = c, (0, 1))\}$. The reservation value depicted in 8(a) is for the first agent (A_1). As can be observed from the figure, the greater the number of agents involved, the lower the reservation value used by the agent. This result is explained by the fact that the greater the number of agents that can potentially affect the effective value, the more reluctant the agent is to extend its exploration in an effort to improve the value it individually finds. Consequently, the average expected benefit decreases as k increases, as illustrated in 8(b).

Figure 9 illustrates the effect of the difference between the agents' exploration costs on their individual and joint expected benefit, in a setting $\{(c_1, (0, 1)), (c_2, (0, 1))\}$. The difference between the exploration costs is captured by the ratios c_1/c_2 and c_2/c_1, keeping the denominator fixed. Therefore, the intersection point between parts (a) and (b) of the figure is in the value 1 over the horizontal axis (i.e., $c_1/c_2 = c_2/c_1 = 1$). As expected, the increase in exploration costs is associated with a decrease in the expected benefit of both agents (as performance is also influenced by the reduction in the reservation value used by the other agent). The most interesting observation from Figure 9 is that the increase in c_1 substantially affects both agents' performance, while the increase in c_2 affects mostly A_2 and only has a slight effect on A_1.

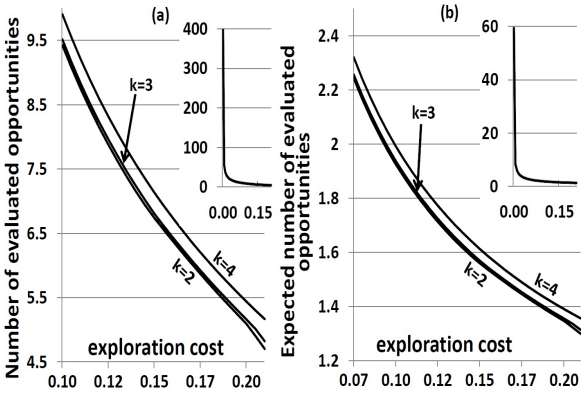

Fig. 10. (a) Maximum number of opportunities that agent A_1 will request to explore in 99.9% of the cases (the 99.9^{th} percentile) as a function of the exploration cost for different k values. (b) The expected number of opportunities that agent A_1 will request to explore as a function of the exploration cost for different k values.

Finally, Figure 10 depicts the number of opportunities the agents choose to obtain as a function of the exploration cost, in a setting where all agents' distribution functions are defined over $(0, 1)$ for different amounts of agents. Figure 10(a) depicts the 99.9^{th} percentile (i.e., the maximum number of opportunities that will need to be evaluated in 99.9% of the cases).

The second depicts the expected number of opportunities obtained. From the two figures, we observe that for most reasonable values of c, the number of opportunities that will need to be evaluated is relatively moderate. The importance of this observation is twofold: first, it shows that even when the number of opportunities available is quite moderate, the probability that an agent will request to evaluate more than those available (if allowing the agent to evaluate as many opportunities as they request) is, in most settings, negligible. Second, it reassures that the sequential exploration is applicable latency-wise, as the number of opportunities evaluated is tightly correlated with the overall latency of the process.

3.3. Comparison to the "naive" strategy

Based on the analysis given above, we can now compare the expected-benefit maximizing strategy for settings of sequential multi-agent exploration with value restrictions with the "naive" one (i.e., the one that does not take into consideration the value restriction resulting from the exploration of former and consequent agents).

One prominent difference between the two is in the exploration sequence. While both strategies rely on assigning a reservation value for each opportunity, the calculation of the reservation values is different in both cases, resulting in a different exploration sequence. Consider, for example, the case of two agents: A_1 and A_2. Agent A_1 can explore 2 opportunities: opportunity o_1^1 with a uniform distribution of values over the interval $(0, 3)$ and the exploration cost is $c_1^1 = 0.65$, and opportunity o_1^2 with a uniform distribution of values over the interval $(0, 1)$ and exploration cost $c_1^2 = 0.01$. Agent A_2 can explore only one opportunity, o_2^1, whose value is derived from a uniform distribution, defined over the interval $(0, 1)$, and its exploration cost is $c_2^1 = 0.01$. The reservation values for o_1^1 and o_1^2 when using the "naive" exploration strategy (according to (1)) are $r_1^1 = 1.03$ and $r_2^1 = 0.86$. Therefore, opportunity o_1^1 ought to be explored first, and only if the value found is lower than 0.86 will opportunity o_1^2 be explored. In contrast, when using the sequential multi-agent exploration with a value restriction strategy (calculated according to Theorem 2), the reservation values to be used are $r_1^1 = -0.18$ and $r_1^2 = 0.69$. Therefore, the sequence in this case is different from the "naive" sequence: opportunity o_1^2 ought to be explored and o_1^1 is never explored.

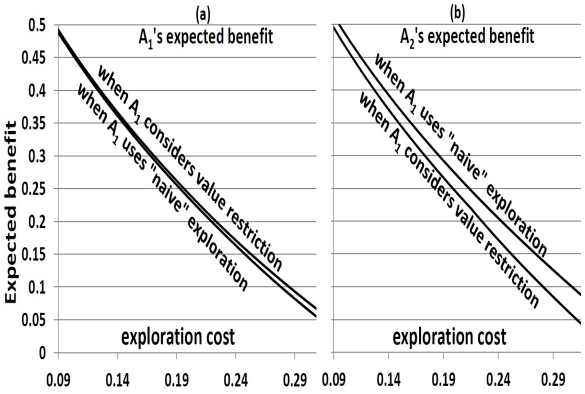

Fig. 11. The expected benefit as a function of the exploration cost when A_2 uses the expected-benefit maximizing strategy with value restriction and A_1 uses either the "naive" strategy or the the expected-benefit maximizing strategy with value restriction: (a)A_1's expected benefit; (b)A_2's expected benefit.

Figure 11 depicts the expected benefit of agent A_1 and agent A_2 as a function of the exploration cost, in a setting $\{(c_1 = c, (0, 1)), (c_2 = c, (0, 1))\}$, when A_2 always uses the expected-benefit maximizing exploration strategy based on the value it receives from A_1, distinguishing between the case where A_1 uses "naive" strategy and when using the expected-benefit maximizing strategy under value restriction. As expected, the expected benefit of agent A_1 increases when it switches from the "naive" strategy to the expected-benefit maximizing strategy according to Theorem 2. The expected benefit of agent A_2 in this case decreases as a result of the change in A_1's strategy. While this may seem intuitive (since if A_1 realizes the values it finds are constrained by the best value found by A_2 it bounds its extent of exploration), this is not true in general. We demonstrate this with the following example: consider a setting of two agents with numerous opportunities. Opportunities available to A_1 yield the values 100 and 110 with an equal probability, and those available to A_2 yield 90 and 110 with an equal probability. The exploration cost is 5 for both agents. In this case, when A_1 is not constrained by the exploration of A_2, it sets a reservation value of 100 and thus its expected benefit is also 100. The reservation values set by A_2 in this case are $r_2 = 90$ when receiving a value $v_1 = 100$ from A_1 and $r_2 = 100$ when receiving 110. The expected benefit of A_2 when A_1 uses the "naive" strategy is thus 95. When A_1 switches to the expected-benefit maximizing strategy, it sets its reservation value to $r_1 = 100$, thus the value transferred to A_2 is always $v_1 = 110$. The expected benefit of A_2

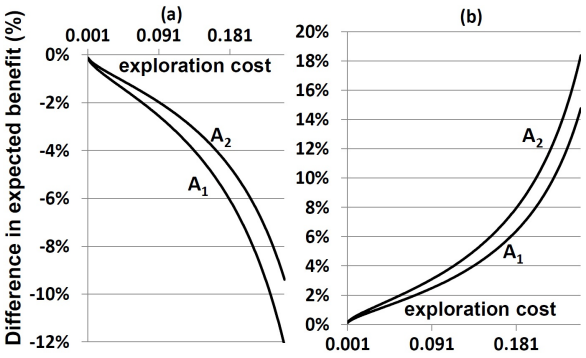

Fig. 12. Differences (in percentages) between the agents' expected benefit under sequential multi-agent exploration and alternative exploration: (a) when the agents follow the "naive" strategy simultaneously; (b) when the agents follow the equilibrium parallel strategy.

in this case is also 100, which is greater than the 95 it would have gained if A_1 had used the "naive" strategy.

The differences, in percentages, between the expected benefit of the agents when using the "naive" reservation values according to [52] and the sequential multi-agent exploration strategy (constraining the values obtained in both cases) as a function of the exploration cost are given in Figure 12(a) for the setting $\{(c_1 = c, (0, 1)), (c_2 = c, (0, 1))\}$. As can be observed from the graph, the use of the sequential multi-agent exploration strategy actually worsened the agents' expected benefit compared to the use of the "naive" exploration strategies. While this may seem surprising, one should keep in mind that the "naive" set of strategies calculated by [52] is not in equilibrium, and each agent has an incentive to deviate from it. This is analyzed in Subsection 3.4. A more suitable comparison in this case is to multi-agent exploration using different strategies that are in equilibrium. An alternative set of exploration strategies that is in equilibrium is the one where all agents explore in parallel, without getting acquainted with the value individually found by the exploration of any of the agents. Consider the case of two agents that are not limited in the number of homogeneous opportunities they can explore. Since the agents do not receive any new information along their (parallel) exploration, their expected-benefit maximizing exploration strategy is reservation value based.

The reservation value in this case can be extracted by solving the following Equations 7 and 8 for r_1 and r_2:

$$c_1 = \int_{y=r_1}^{\infty} f_1(y)(\int_{z=r_2}^{\infty}(\min(y,z) - r_1)f_2(z)dz)dy \quad (7)$$

$$c_2 = \int_{z=r_2}^{\infty} f_2(z)(\int_{y=r_1}^{\infty}(\min(y,z) - r_2)f_1(y)dy)dz \quad (8)$$

Figure 12(b) depicts the difference in the expected benefit of the agents when using the sequential multi-agent exploration with value restriction strategies and when using the equilibrium strategy for the case of simultaneous exploration (in percentages, as a function of the cost used, for the same setting as in Figure 12(a)). As can be observed from Figure 12(b), the sequential multi-agent exploration strategy has the potential to substantially improve both agents' expected benefit.

3.4. Cooperative Behavior and Defection

In the above analysis, all agents were assumed to be self-interested, i.e., each attempted to maximize its own expected benefit. Nevertheless, in various real-life settings, the agents may be cooperative, looking to maximize the sum of their expected benefits, i.e., $\sum_{i=1}^{k}(E[v^*] - EC_i[cost])$. Naturally, the exploration strategies that maximize the latter are different from those used for the self-interested case. Furthermore, while the expected overall joint benefit increases when all agents explore cooperatively, there is often an incentive for some agents to deviate from the cooperative strategy in order to improve their individual expected benefit. In the following paragraphs we present the expected-benefit maximizing exploration strategies to be used in a fully cooperative setting and discuss the dynamics that occur when either of the agents defect from the cooperative strategy.

3.4.1. Fully Cooperative Setting

In the cooperative setting an agent reasoning about exploring an opportunity should consider not only the marginal benefit from such exploration to itself, but also the benefit that all other agents potentially gain from the possible increase in the joint value v^*. In this case we can prove that the joint expected-benefit maximizing strategy is to have all agents use a reservation value based exploration strategy, though with reservation values different from those used for the self-interested case. The proof is identical to the one given for Theorems 1 and 2, where the only change required is the multiplication of the increase in the value v^* by k (more simply put, the value of c_i^j should be divided by k). The reservation value r_k^j of A_k in this case satisfies:

$$c_k^j = \int_{y=r_k^j}^{\infty} k(\min(y, v_{k-1}^*) - r_k^j)f_k^j(y)dy \quad (9)$$

and r_i^j of each agent A_i, where $1 \leq i < k$, satisfies:

$$c_i^j = \int_{y=r_i^j}^{\infty} k(E_{i+1}[v^*|\min(v_{i-1}^*, y)] - E_{i+1}[v^*|r_i^j])f_i^j(y)dy \quad (10)$$

where: $v_0^* = \infty$

As expected, the agents in the cooperative case are likely to explore more extensively (overall), resulting in a greater expected value v^*. This is formally proved in Proposition 2.

Proposition 2. *Both the individual accumulated costs and the expected value $E[v^*]$ at the end of the exploration process, in the fully cooperative case, are greater than those resulting from the self-interested case. Overall, the expected difference among the two components (i.e., sum of expected values minus accumulated costs) is greater in the cooperative case.*

Proof. It is easy to see from (2-3) and (9-10) that the reservation value r_i^j for any value v_{i-1}^* obtained by A_{i-1} (where $1 < i \leq k$) is greater in the cooperative case. Therefore agent A_i, upon receiving a value v_{i-1}^*, will necessarily explore more than in the self-interested case (incurring a greater cost) and its exploration will result in finding a greater expected value. Consequently, agent A_1 receives greater values for each value with which it terminates its exploration, and according to (10) its reservation value r_j^1 necessarily increases. The increase in r_i^j suggests a longer exploration process, i.e., greater exploration costs. Since A_i receives higher values with increased probability, and r_i^j increases as v_{i-1}^* increases, A_i ends up exploring more, overall, and terminates its exploration process with a greater expected value in comparison to the self-interested case. Finally, the joint expected benefit in the cooperative case is greater simply because the agents attempt to directly maximize the sum $\sum_{i=1}^{k}(E[v^*] - EC_i[cost])$ rather than separately maximizing each of its parts. □

Furthermore, even under the permissive assumption that the "naive" reservation values are used, despite not being in equilibrium, the cooperative sequential multi-agent exploration method may substantially improve the joint performance. This is illustrated in Figure 13 for the setting $\{(c_1 = c, (0,1)), (c_2 = c, (0,1))\}$. Here, the expected joint (aggregate) benefit, when using the expected-benefit maximizing with value restriction self-interested strategies, is greater than in the case where the "naive" strategies are applied and smaller than the case where the cooperative strategies are applied.

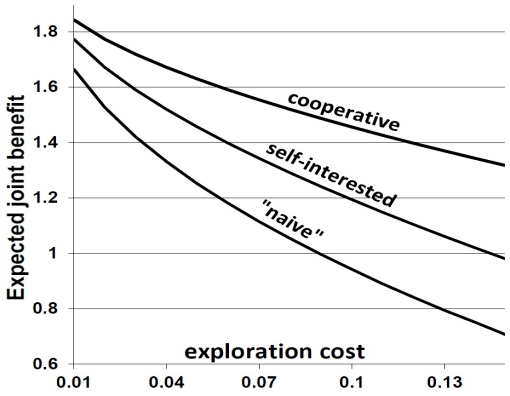

Fig. 13. The expected joint benefit as a function of the exploration cost c, when using the sequential multi-agent strategy with value restriction while the agents are self-interested, when cooperative and when using the "naive" exploration strategy. The setting used is $\{(c_1 = c, (0,1)), (c_2 = c, (0,1))\}$.

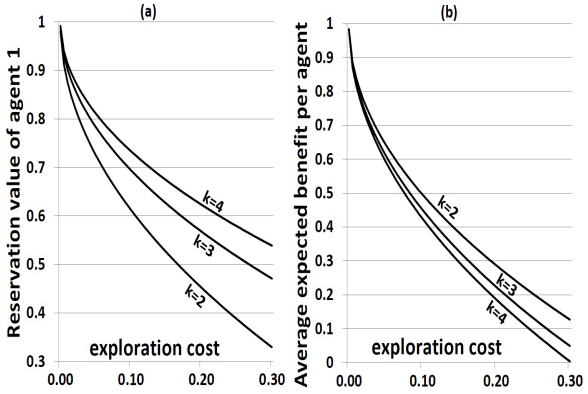

Fig. 14. (a) The optimal reservation values of agent A_1 as a function of the exploration cost c for different amounts of cooperative agents. (b) The average expected benefit (per agent) as a function of the exploration cost c for different amounts of cooperative agents. The setting used is $\{(c_1 = c, (0,1)), .., (c_4 = c, (0,1))\}$.

It is notable that the joint expected benefit does not necessarily improve in comparison to the use of the set of "naive" strategies when using the cooperative new method. For example, when using the reverse setting $\{(c_1 = c, (0,2)), (c_2 = c, (0,1))\}$, the joint expected cooperative benefit is 1.11, while the use of the "naive" strategies yields 1.69. Still, the "naive" set of strategies will never hold in equilibrium.

Figure 14 is the cooperative equivalent of Figure 8. It illustrates the expected joint benefit and the reserva-

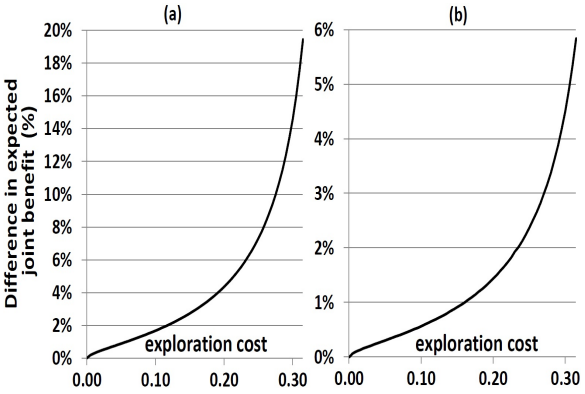

Fig. 15. Differences (in percentages) between the agents' expected joint benefit when using sequential multi-agent exploration and alternative exploration: (a) when the agents follow the "naive" strategy simultaneously; (b) when the agents follow the equilibrium parallel strategy.

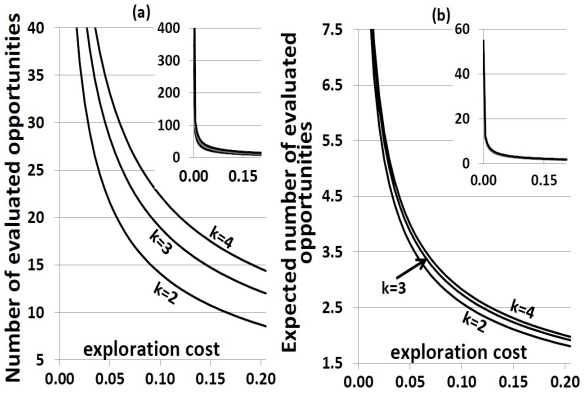

Fig. 16. (a) The maximum number of opportunities that agent A_1 will request to explore in 99.9% of the cases (the 99.9^{th} percentile) as a function of the exploration cost for different k values. (b) The expected number of opportunities that agent A_1 will request to explore as a function of the exploration cost for different k values.

tion value used by A_1 as a function of the exploration cost when all agents use the cooperative strategy, in the setting $\{(c_1 = c, (0, 1)), .., (c_4 = c, (0, 1))\}$. Each curve is correlated with a different number of agents. As expected, as the number of agents that participate in the exploration process increases, the expected joint benefit decreases (14(b)). This is explained by the fact that as the number of agents increases, more exploration needs to take place in order for all agents to obtain a desired value and thus the expected effective value decreases.

The reservation value used by A_1, on the other hand, decreases as the number of agents increases (14(a)). This is in contrast to the behavior observed for the self-interested case (in Figure 8(a)). This is explained by the fact that in this example the k, when using Equation 10, has more influence over r_i^j than the decrease in $E_{i+1}[v^*|\min(v_{i-1}^*, y)]$.

Figure 15 is the cooperative equivalent of Figure 12 (using the same setting), illustrating the difference in percentages between the joint expected benefit when using the sequential multi-agent with value restriction exploration strategy as opposed to using the "naive" strategies (15(a)) and using the cooperative parallel strategies (15(b)). Here, in contrast to the self-interested case, the sequential multi-agent with value restriction strategy improves the performance, not only in comparison to the parallel case, but also in comparison to the "naive" case.

Finally we present Figure 16, which is the cooperative equivalent of Figure 10 (using the same setting), illustrating the number of opportunities that agent A_1 requests to explore as a function of the exploration cost, for different numbers of agents. Figure 16(a) depicts the 99.9^{th} percentile while 16(b) depicts the expected number of opportunities obtained. Here again we observe that for most reasonable values of c, the number of opportunities that will need to be evaluated is relatively moderate.

3.4.2. Incentives to Defect from Cooperation

The cooperative strategies are beneficial when the cooperation can be enforced or when the agents are obligated to the same goal (e.g., working for the same user or users from the same organization). When the cooperation cannot be guaranteed, it will never hold and the agents will use reservation values different from those derived from Equations (9) and (10). For example, regardless of the strategy used by the former $i-1$ agents, agent i can use a reservation value $r_j^{i,\text{defect}}$ according to (2) and (3), rather than $r_j^{i,\text{cooperative}}$ according to (10) and (9), as this strategy maximizes its expected benefit for any value v_{i-1}^* received. If agent A_i uses its self-interested strategy, then we should distinguish between the case where the rest of the agents believe that A_i is cooperative and when they believe it is not cooperative. Naturally the number of possible defecting scenarios is combinatorial.

Figure 17 describes the expected individual and joint benefit as a function of the exploration cost used by the agents for the different variations of cooperation compared to the set of self-interested strategies. The setting used in the figure is $\{(c_1 = c, (0, 1)), (c_2 = c, (0, 1))\}$. One of the curves on each graph represents the case in which both agents use the fully cooperative strategy. Two other curves represent the case in

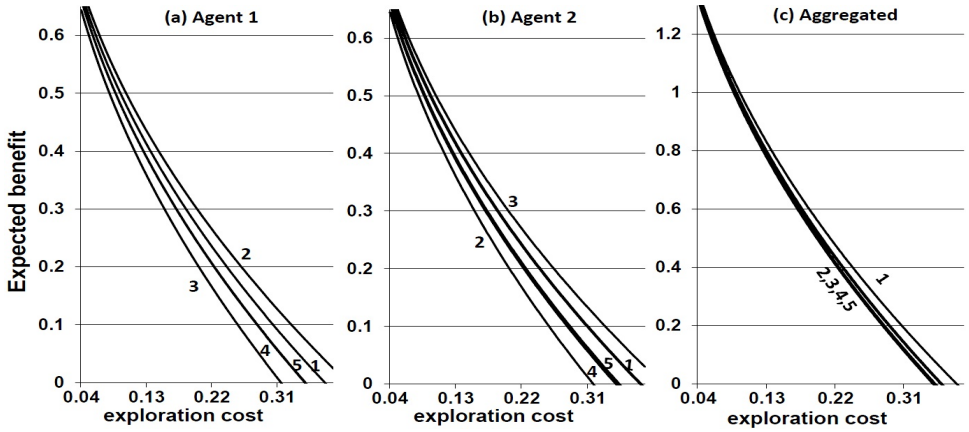

Fig. 17. The case of using the cooperative strategy versus non-cooperative variations: (a) Expected individual benefit for A_1; (b) Expected individual benefit for A_2; and (c) expected joint benefit. The variations used are: (1) A_1 and A_2 are cooperative. (2) A_2 is cooperative and believes A_1 is cooperative, however A_1 is non-cooperative. (3) A_1 is cooperative and believes A_2 is cooperative, however A_2 is non-cooperative. (4) A_1 and A_2 are non-cooperative, however they believe that the other one is cooperative. (5) A_1 and A_2 are self-interested. The setting used is: $\{(c_1 = c, (0, 1)), (c_2 = c, (0, 1))\}$.

which one of the agents uses the cooperative strategy, while the other is self-interested and takes advantage of the fact that the first is being cooperative. The fourth curve represents the case where both agents are self-interested. Finally, the last curve represents the case where both agents are self-interested, though each of them believes the other agent is cooperative. As expected, each agent benefits the most from acting non-cooperatively while the other agent is acting cooperatively (and vice versa, each agent suffers the most when acting cooperatively while the other acts non-cooperatively). Nevertheless, the joint expected-benefit is maximized when both agents are cooperative. In the latter case, the joint expected benefit is substantially better compared to any of the other cases. The case where both agents defect from cooperation is associated with a decreased expected benefit for both agents (compared to acting cooperatively or using the self-interested strategy while assuming the other agent does the same), though it is not as bad as when only one agent defects from cooperation.

3.5. The Use of Side-Payments

As observed in Figure 17 and discussed in former paragraphs, the cooperative set of strategies produces a greater joint benefit. However, it is not stable if the agents are self-interested.

Still, if side-payments are allowed, some agents may benefit from offering other agents to deviate from their expected-benefit maximizing strategy to a different one, more beneficial for the first, and compensate them

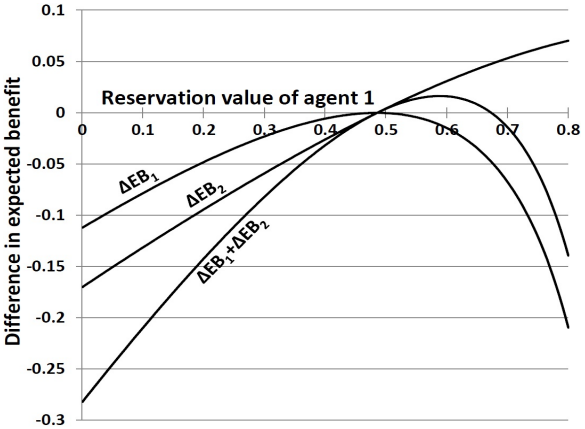

Fig. 18. The difference in the agents and in the joint expected benefit as a function of the reservation value used by A_1 rather than its expected-benefit maximizing strategy's reservation value. The setting used is $\{(c_1 = 0.1, (0, 1)), (c_2 = 0.1, (0, 1))\}$.

for their expected losses. An example of such a case is given in Figure 18, depicting the difference in the expected benefit of each agent and in the joint expected benefit given the reservation value used by A_1 (the horizontal axis) rather than its expected-benefit maximizing strategy's reservation value. The setting used in the example is $\{(c_1 = 0.1, (0, 1)), (c_2 = 0.1, (0, 1))\}$. The expected-benefit maximizing strategy's reservation value for agent A_1 is $r_1 = 0.48$ (in which case all three curves reach zero). As can be observed from the figure, if A_1 deviates from using $r_1 = 0.48$ to any other value from the interval $(0.48, 0.67)$, the improvement achieved in A_2's expected benefit is greater

than the decrease in A_1's expected benefit. Therefore, in any of these cases, A_2 can fully compensate A_1 for deviating to the new reservation value while keeping a positive surplus.

4. Related Work

In many multi-agent environments, autonomous agents may benefit from cooperating and coordinating their actions. Cooperation is mainly useful when an agent is incapable of completing a task by itself or when operating as a group can improve the overall performance [28]. Consequently, group-based cooperative behavior has been suggested in various domains [49,10,50,54,47,18]. The recognition of the advantages encapsulated in teamwork and cooperative behaviors is the main driving force of many coalition formation models in the area of cooperative game theory and MAS [29,45,9,2]. Overall, the majority of cooperation and coalition formation MAS-related research tends to focus on the way coalitions are formed, and consequently concerns issues such as the optimal division of agents into disjoint exhaustive coalitions [43,53], division of coalition payoffs [53] and enforcement methods for interaction protocols [34]. Only a few authors have considered the problem of determining the strategy of a group once formed [22], and no work to date considers exploration strategies for a cooperative exploration of the nature described in this paper.

The problem of an agent engaged in exploration in a costly environment, seeking to maximize its long-term utility, is widely addressed in classical economic search theory (e.g., [40,30,32] and references therein). Over the years, several attempts have been made to adopt search theory concepts for agent-based electronic trading environments associated with exploration costs [7,26]. Despite the richness of search theory and its implications, most models introduced to date have focused on the problem of a single agent that attempts to maximize its own expected benefit. Few studies have attempted to extend the exploration problem beyond a single goal, e.g., attempting to purchase several commodities while facing imperfect information concerning prices [16,6,4]. Some even considered multi-agent cooperative exploration for multiple goals [44,31]. However, none of these works applies any constraints on the values obtained along the exploration process. The only constraint on the values obtained by an agent that can be found in a related work in this area is the availability of recall (i.e., the ability to exploit formerly explored opportunities) [6,32]. To date, to the best of our knowledge, a model of a multi-agent exploration in which one agent's exploration process is constrained by the findings of other agents, as in the cognitive radio application, has not been introduced in this research domain.

Multi-agent exploration can also be found in "two-sided" models (where dual exploration activities are modeled) [46,5,33]. The exploration in these models is used solely for the matching process between the different agents, i.e., for forming appropriate stable partnerships. The value of each agent from a given partnership depends on the partnership itself (e.g., the characteristics of the other agent with whom it partners). In our model, however, the partnership is given a priori and the value of the partnership is derived from an external exploration process performed independently by each agent.

From the Dynamic Spectrum Access application point of view, various spectrum sensing approaches have been proposed, including a cooperative sensing scheme based on distributed detection theory [15], an adaptive MAC layer spectrum sensing [8] and a practical sensing technique that was evaluated in a testbed [38]. Several papers used game theory notions to compare the cooperative and non-cooperative behavior of spectrum sensing and sharing (e.g., [23,25] and references therein). In particular, [20] proposes a scheme in which users exchange "price" signals that indicate the negative effect of interference at the receivers, [21,36] deal with cases in which operators compete for customers as well as portions of available spectrum and [3] analyzes the power control and channel selection problem as a Stackelberg game. Moreover, [24] studies a dynamic spectrum leasing paradigm and [37] proposes a distributed approach, where devices negotiate local channel assignments aiming for a global optimum. Finally, [41] models collaborative spectrum sensing as a nontransferable coalitional game and proposes distributed algorithms for coalition formation, and [51,42] studies the collaborative sensing problem using an evolutionary game framework. Unlike our approach, most of the previous work in the area of Dynamic Spectrum Access and Cognitive Radio focuses on scenarios in which the SUs are in the same geographic area, sense the same set of channels and try to either agree on the same channel or on a set of non-interfering channels. To the best of our knowledge, *searching for channels such that the overall perfor-*

mance is tied to the worst channel selected has not been studied before.

5. Discussion and Conclusions

The sequential multi-agent exploration model extends the traditional exploration models to the case where the process involves several agents that need to engage in individual exploration, and the value of each agent from the process depends on the minimum value found. As discussed throughout the introduction, such a setting arises in various real-life applications and particularly in Dynamic Spectrum Access Systems. The analysis given in this paper proves that the expected-benefit maximizing set of strategies to be used by the agents when using a sequential multi-agent exploration protocol is reservation value based. While this property aligns with a single agent exploration strategy, the equilibrium set of reservation values in the new model are different from those that ought to be used for the single agent case. This also implies that the sequence according to which the different opportunities are explored is often different from the one used in the single agent case. Moreover, a strategy derived according to the latter model can never be in equilibrium, as the remaining agents always have an incentive to use a reservation value lower than the value they obtain. This should thus be taken into considerations by any of the agents.

The sequential nature of the exploration process used enables some level of separation in the analysis: each agent's expected-benefit maximizing strategy is found as a function of the minimum value obtained by former exploring agents and the exploration strategy that will be used by the remaining agents along the sequence. This enables calculation of the equilibrium strategies by solving the resulting Stackelberg game.

While the cooperative setting is highly favorable, it is applicable only when the agents have a joint goal (e.g., when considering family members). In other settings, this set of strategies is not stable, and, as expected, the worst expected joint benefit is obtained when each agent operates self-interestedly while believing that the other agents are cooperative. Another important observation is the substantial effect of the order in which the sequential multi-agent exploration process takes place over the individual and joint benefit. While this issue was left beyond the scope of the analysis, we believe that appropriate methods can be suggested for the agents to negotiate over the order in which they will explore (and possibly come up with schemes for alternating orders in repeated settings) in order to improve the joint and individual expected benefit. Furthermore, the use of side-payments within this context can result in substantial benefits as illustrated in the analysis section.

Additional directions for future research include the development of other multi-agent exploration model variants, e.g., operating simultaneously (as used to a limited extent for illustration purposes), exchanging information throughout the exploration process and even re-initiating exploration by each agent based on the findings received from the other agents. Finally, applying the results to Dynamic Spectrum Access Networks will require taking into account several realistic considerations. These include the exchange of channel quality information between the SUs, the possible operation of a few interfering SUs in the same area (all searching for available channels) and channel mobility resulting from the arrival of PUs claiming back-channels used by the SUs.

Acknowledgments

This work was partially supported by ISF/BSF grants 1401/09 and 2008-404, and the Israeli Ministry of Industry and Trade under project RESCUE.

References

[1] I. F. Akyildiz, W.-Y. Lee, M. C. Vuran, and S. Mohanty. Next generation/dynamic spectrum access/cognitive radio wireless networks: A survey. *Computer Networks*, 50(13):2127–2159, 2006.

[2] S. Albayrak and D. Milosevic. Multi-domain strategy coordination approach for optimal resource usage in agent based filtering framework. *Web Intelligence and Agent Systems*, 4(2):239–253, 2006.

[3] M. Bloem, T. Alpcan, and T. Basar. A stackelberg game for power control and channel allocation in cognitive radio networks. In *Proceedings of the 2nd international conference on Performance evaluation methodologies and tools (ValueTools 2007)*, pages 4:1–4:9, 2007.

[4] K. Burdett and D. A. Malueg. The theory of search for several goods. *Journal of Economic Theory*, 24(3):362–376, 1981.

[5] K. Burdett and R. Wright. Two-sided search with nontransferable utility. *Review of Economic Dynamics*, 1(1):220–245, 1998.

[6] J. A. Carlson and R. P. McAfee. Joint search for several goods. *Journal of Economic Theory*, 32(2):337–345, 1984.

[7] S. P. Choi and J. Liu. Optimal time-constrained trading strategies for autonomous agents. In *Proceedings of the International ICSC Symposium on Multi-Agents and Mobile Agents in*

Virtual Organizations and E-Commerce (MAMA-2000), pages 11–13, 2000.

[8] C.-T. Chou, S. N. Sai, H. Kim, and K. G. Shin. What and how much to gain by spectrum agility? *IEEE Journal on Selected Areas in Communications*, 25(3):576–588, 2007.

[9] E. Crawford and M. M. Veloso. Mechanism design for multi-agent meeting scheduling. *Web Intelligence and Agent Systems*, 4(2):209–220, 2006.

[10] M. B. Dias and T. Sandholm. *TraderBots: A New Paradigm for Robust and Efficient Multirobot Coordination in Dynamic Environments*. PhD thesis, Robotics Institute, Carnegie Mellon University, 2004.

[11] FCC. ET Docket No. 03-222, Notice of Proposed Rule Making and Order, 2003.

[12] FCC. ET Docket No. 04-186, ET Docket No. 02-380, Second Report And Order And Memorandum Opinion And Order, FCC 08-260, 2008.

[13] M. D. Felice, K. R. Chowdhury, and L. Bononi. Analyzing the potential of cooperative cognitive radio technology on inter-vehicle communication. In *Proceedings of the 3rd IFIP Wireless Days Conference 2010*, pages 1–6, 2010.

[14] S. Gal, M. Landsberger, and B. Levykson. A compound strategy for search in the labor market. *International Economic Review*, 22(3):597–608, 1981.

[15] M. Gandetto and C. S. Regazzoni. Spectrum sensing: A distributed approach for cognitive terminals. *IEEE Journal on Selected Areas in Communications*, 25(3):546–557, 2007.

[16] J. Gatti. Multi-commodity consumer search. *Journal of Economic Theory*, 86(2):219–244, 1999.

[17] A. Ghasemi and E.S. Sousa. Spectrum Sensing in Cognitive Radio Networks: Requirements, Challenges and Design Trade-offs. *IEEE Communications Magazine*, 46(4):32–39, 2008.

[18] C. Guttmann, M. P. Georgeff, and I. Rahwan. Collective iterative allocation: Enabling fast and optimal group decision making. *Web Intelligence and Agent Systems*, 8(1):1–35, 2010.

[19] N. Hazon, Y. Aumann, and S. Kraus. Collaborative multi agent physical search with probabilistic knowledge. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI 2009)*, pages 167–174, 2009.

[20] J. Huang, R. A. Berry, and M. L. Honig. Spectrum sharing with distributed interference compensation. In *Proceedings of IEEE DySPANŠ05*, pages 88–93, 2005.

[21] O. Ileri. Demand responsive pricing and competitive spectrum allocation via a spectrum server. In *Proceedings of IEEE DySPANŠ05*, pages 194–202, 2005.

[22] T. Ito, H. Ochi, and T. Shintani. A group-buy protocol based on coalition formation for agent-mediated e-commerce. *IJCIS*, 3(1):11–20, 2002.

[23] K. P. Jagannathan, I. Menache, G. Zussman, and E. Modiano. Non-cooperative spectrum access: the dedicated vs. free spectrum choice. In *Proceedings of the 12th ACM Interational Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 2011)*, page 10, 2011.

[24] S. K. Jayaweera, G. Vazquez-Vilar, and C. Mosquera. Dynamic spectrum leasing: A new paradigm for spectrum sharing in cognitive radio networks. *IEEE Transactions on Vehicular Technology*, 59(5):2328–2339, 2010.

[25] Z. Ji and K.J. Liu. Cognitive radios for dynamic spectrum access - dynamic spectrum sharing: A game theoretical overview. *IEEE Communications Magazine*, 45(5):88–94, 2007.

[26] J. O. Kephart and A. Greenwald. Shopbot economics. *Autonomous Agents and Multi-Agent Systems*, 5(3):255–287, 2002.

[27] J. O. Kephart, J. E. Hanson, and A. Greenwald. Dynamic pricing by software agents. *Computer Networks*, 32(6):731–752, 2000.

[28] K. Lerman and O. Shehory. Coalition formation for large-scale electronic markets. In *Proceedings of the 4th International Conference on Multi-Agent Systems (ICMAS 2000)*, pages 167–174, 2000.

[29] C. Li, U. Rajan, S. Chawla, and K. Sycara-Cyranski. Mechanisms for coalition formation and cost sharing in an electronic marketplace. In *Proceedings of the 5th International Conference on Electronic Commerce (ICEC 2003)*, pages 68–77, 2003.

[30] S. Lippman and J. McCall. The economics of job search: A survey. *Economic Inquiry*, 14(3):347–368, 1976.

[31] E. Manisterski, D. Sarne, and S. Kraus. Cooperative search with concurrent interactions. *Artificial Intelligence Research*, 32:1–36, 2008.

[32] J. McMillan and M. Rothschild. Search. In *Proceedings of Handbook of Game Theory with Economic Applications*, pages 905–927, 1994.

[33] J. M. McNamara and E. J. Collins. The job search problem as an employer-candidate game. *Journal of Applied Probability*, 27(4):815–827, 1990.

[34] P. Michiardi and R. Molva. Analysis of coalition formation and cooperation strategies in mobile ad hoc networks. *Ad Hoc Networks*, 3:193–219, 2005.

[35] J. Mitola III. Cognitive Radio for Flexible Mobile Multimedia Communications. *MONET*, 6(5):435–441, 2001.

[36] D. Niyato and E. Hossain. Competitive pricing for spectrum sharing in cognitive radio networks: Dynamic game, inefficiency of nash equilibrium, and collusion. *IEEE Journal on Selected Areas in Communications*, 26(1):192–202, 2008.

[37] C. Peng, H. Zheng, and B. Y. Zhao. Utilization and fairness in spectrum assignment for opportunistic spectrum access. *MONET*, 11(4):555–576, 2006.

[38] H. Rahul, N. Kushman, D. Katabi, C. Sodini, and F. Edalat. Learning to share: narrowband-friendly wideband wireless networks. In *Proceedings of the ACM SIGCOMM 2008 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, pages 147–158, 2008.

[39] I. Rochlin, D. Sarne, and G. Zussman. Sequential multilateral search for a common goal. In *Proceedings of the 2011 IEEE/WIC/ACM International Conference on Intelligent Agent Technology (IAT 2011)*, pages 349–356, 2011.

[40] M. Rothschild. Searching for the lowest price when the distribution of prices is unknown. *Journal of Political Economy*, 82(4):689–711, 1974.

[41] W. Saad, Z. Han, T. Basar, M. Debbah, and A. Hjorungnes. Coalition formation games for collaborative spectrum sensing. *IEEE Transactions on Vehicular Technology*, 60(1):276 –297, 2011.

[42] W. Saad, Z. Han, M. Debbah, A. Hjorungnes, and T. Basar. Coalitional games for distributed collaborative spectrum sensing in cognitive radio networks. *CoRR*, pages 2114–2122, 2009.

[43] T. Sandholm, K. Larson, M. Andersson, O. Shehory, and F. Tohme. Coalition structure generation with worst case guarantees. *Artificial Intelligence*, 111(1):209–238, 1999.

[44] D. Sarne, E. Manisterski, and S. Kraus. Multi-goal economic search using dynamic search structures. *Autonomous Agents and Multi-Agent Systems*, 21(1-2):204–236, 2010.

[45] O. Shehory and S. Kraus. Methods for task allocation via agent coalition formation. *Artificial Intelligence*, 101(1-2):165–200, 1998.

[46] R. Shimer and L. Smith. Assortative matching and search. *Econometrica*, 68(2):343–370, 2000.

[47] G. Singh and R. Weiskircher. A multi-agent system for decentralised fractional shared resource constraint scheduling. *Web Intelligence and Agent Systems*, 9(2):99–108, 2011.

[48] S. F. Smith, A. Gallagher, and T. L. Zimmerman. Distributed management of flexible times schedules. In *Proceedings of the 6th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2007)*, pages 472–479, 2007.

[49] S. Talukdar, L. Baerentzen, A. Gove, and P. S. de Souza. Asynchronous teams: Cooperation schemes for autonomous agents. *Heuristics*, 4(4):295–321, 1998.

[50] M. Tsvetovat, K. P. Sycara, Y. Chen, and J. Ying. Customer coalitions in electronic markets. In *Proceedings of Agent-Mediated Electronic Commerce III, Current Issues in Agent-Based Electronic Commerce Systems (includes revised papers from AMEC 2000 Workshop)*, pages 121–138, 2000.

[51] B. Wang, K. J. Liu, and T. C. Clancy. Evolutionary cooperative spectrum sensing game: how to collaborate? *IEEE Transactions on Communications*, 58(3):890–900, 2010.

[52] M. L. Weitzman. Optimal search for the best alternative. *Econometrica*, 47(3):641–654, 1979.

[53] J. Yamamoto and K. Sycara. A stable and efficient buyer coalition formation scheme for e-marketplaces. In *Proceedings of the 5th international conference on Autonomous agents (AGENTS '01)*, pages 576–583, 2001.

[54] J. Yen, X. Fan, and R. A. Volz. Information needs in agent teamwork. *Web Intelligence and Agent Systems*, 2(4):231–247, 2004.

Appendices

A. Proofs

A.1. Proof of Theorem 1

Proof. The structure of the proof follows the one given in [52] for the case where no restrictions are made on the value found and augments it to our case.

The reservation value r_k^j set by agent A_k is necessarily smaller than v_{k-1}^* because otherwise no exploration will take place by that agent, and the agent needs to explore at least one opportunity in order for the partnership to be formed. In order to prove that r_k^j, which satisfies (2) always exists, we consider the term $H_k^j(r_k^j) = \int_{y=r_k^j}^{\infty}(\min(y, v_{k-1}^*) - r_k^j)f_k^j(y)dy$. The function $H_k^j(r_k^j)$ is continuous and monotonic decreasing in r_k^j. Since $r_k^j < v_{k-1}^*$, it satisfies: $H_k^j(-\infty) = \infty$, $H_k^j(\infty) = 0$.

For the inductive part, we begin with the case of having a single opportunity. Here, the right-hand side of (2) can be interpreted as the expected additional gain from obtaining the minimum between the value of that opportunity and v_{k-1}^*, if the agent is already guaranteed a value $r_k^j < v_{k-1}^*$. Obtaining the value of the opportunity in this case is thus beneficial only if the expected additional gain is greater than c_k^j. Since H_k^j is monotonic decreasing in r_k^j, the opportunity should be explored whenever the value that can be guaranteed from A_k's exploration so far is less than r_k^j and r_k^j is in fact a reservation value.

Assume that the reservation-value based strategy is optimal for the case of $n'_k < n_k$ unexplored opportunities when the best value found so far by A_k is y. We need to prove that for the case of $n'_k + 1$ unexplored opportunities and best known value y, the expected-benefit maximizing exploration strategy is also reservation-value based and obeys Equation 2. Consider the opportunity o_k^j associated with the highest reservation value (calculated according to Equation 2) among the n'_k+1 unexplored opportunities. In order to reason about exploring that opportunity we distinguish between two cases. The first is where $y \geq r_k^j$ for each opportunity o_k^j which has not been explored yet. The second is where there exists an unexplored opportunity o_k^j for which $y < r_k^j$. In the first case ($y \geq r_k^j$ for each unexplored opportunity o_k^j), if exploring one of the unexplored opportunities, the agent is left with n'_k opportunities whose reservation values are necessarily less than y. Therefore, according to the inductive assumption, the exploration should terminate. The decision in that case of whether or not to explore an opportunity o_k^j should thus be made solely based on the benefit of exploring o_k^j (constrained by v_{k-1}^*) and the cost c_k^j. The value of exploring o_k^j (and terminating the exploration right after) is given by H_k^j. For the case of $y \geq r_k^j$, this cost is necessarily less than c_k^j (since H_k^j is monotonic decreasing in r_k^j) and thus the opportunity should not be explored.

For the case where there exists an unexplored opportunity o_k^j for which $y < r_k^j$, it is guaranteed that at least one opportunity should be explored (since according to Equation 2, exploring o_k^j and terminating the exploration is preferable to not exploring at all). Therefore all that is needed is to prove that first exploring the opportunity associated with the largest reservation value

out of the $n'_k + 1$ unexplored opportunities (according to Equation 2), denoted o^z_k, rather than any other opportunity $o^b_k \neq o^z_k$, yields a better expected benefit.

The expected benefit from obtaining the value of o^z_k and then following the optimal strategy for the remaining n'_k opportunities according to the induction assumption is greater than obtaining the value of any opportunity o^z_k ($i \neq z$) and then following the optimal strategy for the remaining n'_k opportunities according to the induction assumption. Let o^h_k denote the opportunity associated with the second largest reservation value of the $n'_k + 1$ unexplored opportunities.

We define the following two strategies:

- Strategy A - explore opportunity o^z_k. If the value obtained is greater than the reservation value of opportunity o^h_k, then the exploration should be terminated. Otherwise explore opportunity o^b_k. Notice that the expected benefit from this strategy is necessarily less than the expected benefit of the optimal strategy that follows the exploration of o^z_k, since the exploration according to A continues with o^b_k rather than o^h_k according to the inductive assumption.
- Strategy B - explore opportunity $o^b_k \neq o^z_k$. If the value obtained from o^b_k is smaller than r^z_k, then explore o^z_k next (according to the induction assumption for n'_k unexplored opportunities).

In order to prove that the expected benefit of strategy A is greater than the expected benefit of strategy B the following assisting notations are used:

$$\pi_z = prob(x_z \geq r^z_k), w_z = E[\min(x_z, v^*_{k-1}) | x_z \geq r^z_k]$$
$$\pi_b = prob(x_b \geq r^z_k), w_b = E[\min(x_b, v^*_{k-1}) | x_b \geq r^z_k]$$
$$\lambda_z = prob(r^h_k \leq x_z < r^z_k)$$
$$v_z = E[\min(x_z, v^*_{k-1}) | r^h_k \leq x_z < r^z_k]$$
$$\vartheta_z = E[\min(\max(x_z, y), v^*_{k-1}) | r^h_k \leq x_z < r^z_k]$$
$$\lambda_b = prob(r^h_k \leq x_b < r^z_k)$$
$$v_b = E[\min(x_b, v^*_{k-1}) | r^h_k \leq x_b < r^z_k]$$
$$\vartheta_b = E[\min(\max(x_b, y), v^*_{k-1}) | r^h_k \leq x_b < r^z_k]$$
$$\mu_b = prob(r^b_k \leq x_b < r^h_k)$$
$$u_b = E[\min(x_b, v^*_{k-1}) | r^b_k \leq x_b < r^h_k]$$
$$d = E[\min(\max(x_z, x_b, y), v^*_{k-1}) | | r^h_k \leq x_z < r^z_k; r^h_k \leq x_b < r^z_k]$$

Since according to v^*_{k-1} and r^j_k's definition $v^*_{k-1} > r^j_k$, therefore:

$$v_z = E[x_z | r^h_k \leq x_z < r^z_k]$$
$$\vartheta_z = E[max(x_z, y) | r^h_k \leq x_z < r^z_k]$$
$$v_b = E[x_b | r^h_k \leq x_b < r^z_k]$$
$$\vartheta_b = E[max(x_b, y) | r^h_k \leq x_b < r^z_k]$$
$$u_b = E[x_b | r^b_k \leq x_b < r^h_k]$$
$$d = E[max(x_z, x_b, y) | r^h_k \leq x_z < r^z_k; r^h_k \leq x_b < r^z_k]$$

In addition, we use ϕ to denote the expected benefit from resuming the exploration after exploring opportunities o^z_k and o^b_k if the maximum of y, x_z and x_b is less than r^h_k. The expected benefit of strategy A, denoted E_A, is:

$$\begin{aligned} E_A = & -c^z_k + \pi_z w_z + \lambda_z \vartheta_z \\ & + (1 - \pi_z - \lambda_z)[-c^b_k + \pi_b w_b + \lambda_b \vartheta_b] \\ & + (1 - \pi_z - \lambda_z)(1 - \pi_b - \lambda_b)\phi \end{aligned} \tag{11}$$

The expected value of strategy B is:

$$\begin{aligned} E_B = & -c^b_k + \pi_b w_b \\ & + \lambda_b[-c^z_k + \pi_z w_z + \lambda_z d + (1 - \pi_z - \lambda_z)\vartheta_b] \\ & + (1 - \pi_b - \lambda_b)[-c^z_k + \pi_z w_z + \lambda_z \vartheta_z] \\ & + (1 - \pi_b - \lambda_b)(1 - \pi_z - \lambda_z)\phi \end{aligned} \tag{12}$$

Subtracting (12) from (11) obtains:

$$\begin{aligned} E_A - E_B = & (\pi_z + \lambda_z)[c^b_k - \pi_b w_b] \\ & + \pi_b[\pi_z w_z + \lambda_z \vartheta_z - c^z_k] + \lambda_b \lambda_z (\vartheta_z - d) \end{aligned} \tag{13}$$

Now, based on Equation 2, the following should hold:

$$c^z_k = \pi_z(w_z - r^z_k) \tag{14}$$
$$c^b_k = \pi_b(w_b - r^b_k) + \lambda_b(v_b - r^b_k) + \mu_b(u_b - r^b_k)$$

Substituting (13) in (14), we obtain:

$$\begin{aligned} E_A - E_B = & \pi_b \pi_z (r^z_k - r^b_k) + \pi_z \lambda_b (v_b - r^b_k) \\ & + \lambda_z \pi_b (\vartheta_z - r^b_k) + \mu_b (\pi_z + \lambda_z)(u_b - r^b_k) \\ & + \lambda_b \lambda_z (v_b + \vartheta_z - r^b_k - d) \end{aligned} \tag{15}$$

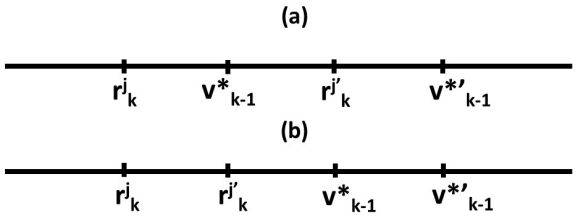

Fig. 19. An illustration of the cases:
(a)$r_k^j < v_{k-1}^* < r_k^{j'} < v_{k-1}^{*'}$;(b)$r_k^j < r_k^{j'} < v_{k-1}^* < v_{k-1}^{*'}$

Notice that by definition $r_k^z > r_k^b$, $v_b > r_k^b$, $\vartheta_z > r_k^b$ and $u_b > r_k^b$. Therefore, in order for $E_A - E_B$ to be positive, the following should hold: $(v_b + \vartheta_z - r_k^b - d) \geq 0$. The latter is obtained by showing that $d \leq v_b + \vartheta_z - r_k^h \leq v_b + \vartheta_z - r_k^b$. According to the definition of d above, the following holds:

$$
\begin{aligned}
&d = r_k^h + E[max(max(x_z, y) - r_k^h, x_b - r_k^h) \\
&| r_k^h \leq x_z < r_k^z; r_k^h \leq x_b < r_k^z] \\
&\leq r_k^h + E[(max(x_z, y) - r_k^h + x_b - r_k^h) \\
&| r_k^h \leq x_z < r_k^z; r_k^h \leq x_b < r_k^z] \\
&= v_b + \vartheta_z - r_k^h \leq v_b + \vartheta_z - r_k^b
\end{aligned}
$$

which proves that $E_A - E_B > 0$, and therefore o_k^z should be explored in that case. □

A.2. Proof for Proposition 1

Proof. We prove that for any two pairs (v_{k-1}^*, r_k^j) and $(v_{k-1}^{*'}, r_k^{j'})$ satisfying Equation 2 and $v_{k-1}^{*'} > v_{k-1}^*$ the relation $v_{k-1}^{*'} - v_{k-1}^* \geq r_k^{j'} - r_k^j$ holds. Assume otherwise, i.e., $v_{k-1}^{*'} - v_{k-1}^* \geq r_k^{j'} - r_k^j$. Since r_k^j increases as v_{k-1}^* increases, $r_k^{j'} > r_k^j$ and thus we only need to consider two cases. The first is when $r_k^j < v_{k-1}^* < r_k^{j'} < v_{k-1}^{*'}$ (Figure 19(a)) and the second is when $r_k^j < r_k^{j'} < v_{k-1}^* < v_{k-1}^{*'}$ (Figure 19(b)).

For the case $r_k^j < v_{k-1}^* < r_k^{j'} < v_{k-1}^{*'}$, the following holds (according to (2)):

$$
c_k^j = \int_{y=r_k^j}^{v_{k-1}^*} (y - r_k^j) f_k^j(y) dy + (v_{k-1}^* - r_k^j) \int_{y=v_{k-1}^*}^{\infty} f_k^j(y) dy \tag{16}
$$

$$
c_k^j = \int_{y=r_k^{j'}}^{v_{k-1}^{*'}} (y - r_k^{j'}) f_k^j(y) dy + (v_{k-1}^{*'} - r_k^{j'}) \int_{y=v_{k-1}^{*'}}^{\infty} f_k^j(y) dy
$$

Notice that:

$$
\int_{y=r_k^{j'}}^{v_{k-1}^{*'}} (y - r_k^{j'}) f_k^j(y) dy + (v_{k-1}^{*'} - r_k^{j'}) \int_{y=v_{k-1}^{*'}}^{\infty} f_k^j(y) dy \tag{17}
$$

$$
< (v_{k-1}^{*'} - r_k^{j'}) \int_{y=r_k^{j'}}^{v_{k-1}^{*'}} f_k^j(y) dy + (v_{k-1}^{*'} - r_k^{j'}) \int_{y=v_{k-1}^{*'}}^{\infty} f_k^j(y) dy
$$

$$
= (v_{k-1}^{*'} - r_k^{j'}) \int_{y=r_k^{j'}}^{\infty} f_k^j(y) dy
$$

and since $r_k^{j'} > r_j^k$:

$$
(v_{k-1}^{*'} - r_k^{j'}) \int_{y=r_k^{j'}}^{\infty} f_k^j(y) dy < (v_{k-1}^* - r_k^j) \int_{y=v_{k-1}^*}^{\infty} f_k^j(y) dy < c_k^j \tag{18}
$$

which contradicts (16). Similarly, for the case $r_k^j < r_k^{j'} < v_{k-1}^* < v_{k-1}^{*'}$, according to (2):

$$
\begin{aligned}
c_k^j &= \int_{y=r_k^j}^{r_k^{j'}} (y - r_k^j) f_k^j(y) dy + \int_{y=r_k^{j'}}^{v_{k-1}^*} y f_k^j(y) dy \\
&- r_k^j \int_{y=r_k^{j'}}^{v_{k-1}^*} f_k^j(y) dy + (v_{k-1}^* - r_k^j) \int_{y=v_{k-1}^*}^{\infty} f_k^j(y) dy
\end{aligned} \tag{19}
$$

$$
\begin{aligned}
c_k^j &= \int_{y=r_k^{j'}}^{v_{k-1}^*} y f_k^j(y) dy - r_k^{j'} \int_{y=r_k^{j'}}^{v_{k-1}^*} f_k^j(y) dy \\
&+ \int_{y=v_{k-1}^*}^{v_{k-1}^{*'}} (y - r_k^{j'}) f_k^j(y) dy + (v_{k-1}^{*'} - r_k^{j'}) \int_{y=v_{k-1}^*}^{\infty} f_k^j(y) dy \\
&= \int_{y=r_k^{j'}}^{v_{k-1}^*} y f_k^j(y) dy - r_k^{j'} \int_{y=r_k^{j'}}^{v_{k-1}^*} f_k^j(y) dy \\
&+ \int_{y=v_{k-1}^*}^{v_{k-1}^{*'}} (y - r_k^{j'}) f_k^j(y) dy + (v_{k-1}^{*'} - r_k^{j'}) \int_{y=v_{k-1}^*}^{\infty} f_k^j(y) dy \\
&- (v_{k-1}^{*'} - r_k^{j'}) \int_{y=v_{k-1}^*}^{v_{k-1}^{*'}} f_k^j(y) dy
\end{aligned} \tag{20}
$$

Subtracting (20) from (19) obtains:

$$0 = \int_{y=r_k^j}^{r_k^{j'}} (y - r_k^j) f_k^j(y) dy + (r_k^{j'} - r_k^j) \int_{y=r_k^{j'}}^{v_{k-1}^*} f_k^j(y) dy \quad (21)$$

$$+ (v_{k-1}^* - r_k^j - (v_{k-1}^{*'} - r_k^{j'})) \int_{y=v_{k-1}^*}^{\infty} f_k^j(y) dy$$

$$+ \int_{y=v_{k-1}^*}^{v_{k-1}^{*'}} (v_{k-1}^{*'} - y) f_k^j(y) dy$$

Since all elements in (21) except for $(v_{k-1}^* - r_k^j - (v_{k-1}^{*'} - r_k^{j'})) \int_{y=v_{k-1}^*}^{\infty} f_k^j(y) dy$ are positive, the latter term must be negative in order for the equation to hold. Therefore $v_{k-1}^* - r_k^j < v_{k-1}^{*'} - r_k^{j'}$, which contradicts the initial assumption.

□

A.3. Proof for Theorem 2

Proof. In order to prove the Theorem, we first prove that the expected effective value increases as the value any of the agents receives along the way increases. While this may seem intuitive, it requires a formal proof as given in Lemma 1.

Lemma 1. $E_{i+1}[v^*|min(v_{i-1}^{*'}, y)] \leq E_{i+1}[v^*|min(v_{i-1}^*, y)]$ *for any* v_{i-1}^* *and* $v_{i-1}^{*'}$ *satisfying* $v_{i-1}^{*'} < v_{i-1}^*$.

Proof. Assume otherwise, i.e., $E_{i+1}[v^*|\min(v_{i-1}^{*'}, y)] < E_{i+1}[v^*|\min(v_{i-1}^*, y)]$. Obviously agent A_i can perform its exploration assuming it had received the value $v_{i-1}^{*'}$ and transfers the value $v_i^* = \min(y, v_{i-1}^{*'})$. In this case the values returned by the remaining $k - i$ agents, $v_{i+1}, ..., v_k$, remain the same as if A_i originally received v_{i-1}^*. However, the effective value in this case is actually $\min(v_{i-1}^*, y, v_{i+1}, ..., v_k)$, as opposed to $\min(v_{i-1}^{*'}, y, v_{i+1}, ..., v_k)$ in the case where A_i received $v_{i-1}^{*'}$. Since $v_{i-1}^{*'} < v_{i-1}^*$, it follows that $E_{i+1}[v^*|\min(v_{i-1}^{*'}, y)] < E_{i+1}[v^*|\min(v_{i-1}^*, y)]$ cannot hold. □

Corollary 1. $E_i[v^{*'}|v_{i-1}^{*'}] \leq E_i[v^*|v_{i-1}^*]$ *for any* v_{i-1}^* *and* $v_{i-1}^{*'}$ *satisfying* $v_{i-1}^{*'} < v_{i-1}^*$.

The above facilitates a proof for Theorem 2 according to the methodology used in the proof of Theorem 1 with the appropriate modifications of expected value calculation.

The relation $r_i^j < v_{i-1}^*$ holds for the same considerations give in the proof for Theorem 1. Similarly, the existence of r_i^j results from the continuity and monotonicity of the term $\int_{y=r_i^j}^{\infty} (E_{i+1}[v^*|\min(v_{i-1}^*, y)] - E_{i+1}[v^*|r_i^j]) f_i^j(y) dy$ in r_i^j and its values for $r_i^j = -\infty$ and $r_i^j = \infty$.

For the case of the single opportunity, the correctness derives from comparing the cost c_i^j with the additional expected gain from moving on to the next agent's exploration process after exploring that opportunity (i.e. $E_{i+1}[v^*|\min(v_{i-1}^*, y)]$) in comparison to the case of not exploring that opportunity (i.e. $E_{i+1}[v^*|r_i^j]$).

For the inductive part of the proof, the analysis of the case of $y \geq r_i^j$ remains unchanged. Similarly, for the case where $y < r_i^j$, the same alternative strategies A and B are defined. This time, in order to prove that the expected benefit of strategy A is greater than the expected benefit of strategy B, the notations' definitions need to be augmented as follows:

$$\pi_z = prob(x_z \geq r_i^z)$$

$$w_z = E[E_{i+1}[v^*|\min(x_z, v_{i-1}^*)]|x_z \geq r_i^z]$$

$$\pi_b = prob(x_b \geq r_i^z)$$

$$w_b = E[E_{i+1}[v^*|\min(x_b, v_{i-1}^*)]|x_b \geq r_i^z]$$

$$\lambda_z = prob(r_i^h \leq x_z < r_i^z)$$

$$v_z = E[E_{i+1}[v^*|\min(x_z, v_{i-1}^*)]|r_i^h \leq x_z < r_i^z]$$

$$\vartheta_z = E[E_{i+1}[v^*|\min(\max(x_z, y), v_{i-1}^*)]|r_i^h \leq x_z < r_i^z]$$

$$\lambda_b = prob(r_i^h \leq x_b < r_i^z)$$

$$v_b = E[E_{i+1}[v^*|\min(x_b, v_{i-1}^*)]|r_i^h \leq x_b < r_i^z]$$

$$\vartheta_b = E[E_{i+1}[v^*|\min(\max(x_b, y), v_{i-1}^*)]|r_i^h \leq x_b < r_i^z]$$

$$\mu_b = prob(r_i^b \leq x_b < r_i^h)$$

$$u_b = E[E_{i+1}[v^*|\min(x_b, v_{i-1}^*)]|r_i^b \leq x_b < r_i^h]$$

$$d = E[E_{i+1}[v^*|\min(\max(x_z, x_b, y), v_{i-1}^*)]$$
$$|r_i^h \leq x_z < r_i^z; r_i^h \leq x_b < r_i^z]$$

Since $v^*_{i-1} > r^j_i$, the following hold:

$$
\begin{aligned}
v_z &= E[E_{i+1}[v^*|x_z]|r^h_i \le x_z < r^z_i] \\
\vartheta_z &= E[E_{i+1}[v^*|\max(x_z, y)]|r^h_i \le x_z < r^z_i] \\
v_b &= E[E_{i+1}[v^*|x_b]|r^h_i \le x_b < r^z_i] \\
\vartheta_b &= E[E_{i+1}[v^*|\max(x_b, y)]|r^h_i \le x_b < r^z_i] \\
u_b &= E[E_{i+1}[v^*|x_b]|r^b_i \le x_b < r^h_i] \\
d{=}&E[E_{i+1}[v^*|\max(x_z, x_b, y)]|r^h_i{\le}x_z{<}r^z_i; r^h_i{\le}x_b{<}r^z_i]
\end{aligned}
$$

The notation ϕ is left unchanged. The expected benefit of strategies A and B, denoted E_A and E_B, respectively, remain unchanged and are given by Equations 11 and 12. Consequently, their subtraction yields the same term given in Equation 13.

Now, based on Equation 3, the following should hold:

$$
\begin{aligned}
c^z_i &= \pi_z(w_z - E_{i+1}[v^*|r^z_i]) \qquad (22) \\
c^b_i &= \pi_b(w_b - E_{i+1}[v^*|r^b_i]) + \lambda_b(v_b - E_{i+1}[v^*|r^b_i]) \\
&\quad + \mu_b(u_b - E_{i+1}[v^*|r^b_i])
\end{aligned}
$$

Substituting (13) in (22), we obtain:

$$
\begin{aligned}
&E_A - E_B = \pi_z\lambda_b(v_b - E_{i+1}[v^*|r^b_i]) \\
&+ \mu_b(\lambda_z + \pi_z)(u_b - E_{i+1}[v^*|r^b_i]) \\
&+ \pi_b\pi_z(E_{i+1}[v^*|r^z_i] - E_{i+1}[v^*|r^b_i]) \\
&+ \pi_b\lambda_z(\vartheta_z - E_{i+1}[v^*|r^b_i]) \\
&+ \lambda_b\lambda_z(\vartheta_z + v_b - E_{i+1}[v^*|r^b_i] - d)
\end{aligned}
$$

Notice that according to Lemma 1, $E_{i+1}[v^*|r^z_i] > E_{i+1}[v^*|r^b_i]$. Furthermore, based on the notations' definitions: $v_b > E_{i+1}[v^*|r^b_i]$, $\vartheta_z > E_{i+1}[v^*|r^b_i]$ and $u_b > E_{i+1}[v^*|r^b_i]$. Therefore, in order for $E_A - E_B$ to be positive, the following should hold: $(v_b + \vartheta_z - E_{i+1}[v^*|r^b_i] - d) \ge 0$. The latter is obtained by showing that $d \le v_b + \vartheta_z - E_{i+1}[v^*|r^h_i] \le v_b + \vartheta_z - E_{i+1}[v^*|r^b_i]$. According to the definition of d above, the following holds:

$$
\begin{aligned}
&d = E[\max(E_{i+1}[v^*|\max(x_z, y)], E_{i+1}[v^*|x_b]) \\
&|r^h_i \le x_z < r^z_i; r^h_i \le x_b < r^z_i] \\
&{=}E_{i+1}[v^*|r^h_i]{+}E[\max(E_{i+1}[v^*|\max(x_z, y)] \\
&{-}E_{i+1}[v^*|r^h_i], E_{i+1}[v^*|x_b]{-}E_{i+1}[v^*|r^h_i]) \\
&|r^h_i \le x_z < r^z_i; r^h_i \le x_b < r^z_i] \\
&\le E_{i+1}[v^*|r^h_i] + E[E_{i+1}[v^*|\max(x_z, y)] \\
&- E_{i+1}[v^*|r^h_i] + E_{i+1}[v^*|x_b] \\
&- E_{i+1}[v^*|r^h_i]|r^h_i \le x_z < r^z_i; r^h_i \le x_b < r^z_i] \\
&= \vartheta_z + v_b - E_{i+1}[v^*|r^h_i] \le v_b + \vartheta_z - E_{i+1}[v^*|r^b_i]
\end{aligned}
$$

which proves that $E_A - E_B > 0$, and therefore o^z_i should be explored in that case. □

B. Nomenclature

Notation	Meaning
k	The number of individual agents that need to establish an ad-hoc partnership
A_i	The i^{th} agent i in the exploration process ($1 \leq i \leq k$)
n_i	Number of opportunities available to agent A_i ($i = 1, \ldots, k$)
O_i	The set of opportunities available to agent A_i ($i = 1, \ldots, k$)
$\bar{O}_i$	The set of opportunities which values have not been obtained yet out of those available to A_i ($\bar{O}_i \subseteq O_i$)
o_i^j	The j^{th} ($1 \leq j \leq n_i$) opportunity available to agent A_i ($i = 1, \ldots, k$)
v_i	The value seen in the partnership by agent A_i
v^*	The minimum among the values seen in the partnership by all agents ("effective value")
v_i^*	The minimum among the values seen in the partnership by the first i agents in the sequence
$f_i^j(y)$, $F_i^j(y)$	The probability density function and cumulative distribution function of opportunity o_i^j ($1 \leq j \leq n_i$) available to agent A_i ($i = 1, \ldots, k$)
c_i^j	The exploration cost of opportunity o_i^j ($1 \leq j \leq n_i$) available to agent A_i
r_i^j	The reservation value that agent A_i assigns to opportunity o_i^j ($1 \leq j \leq n_i$)
$(v_{i-1}^*, w, \bar{O}_i)$	The state of agent A_i, where v_{i-1}^* is the minimum value obtained by the previous $i-1$ agents that have already finished their exploring, w is the best value found so far by A_i and $\bar{O}_i$ is the set of opportunities which values have not been obtained yet
$E_i[v^*\|(v_{i-1}^*, w, \bar{O}_i)]$	The expected effective value if agent A_i is about to start its exploration process, given state $(v_{i-1}^*, w, \bar{O}_i)$
$E_i[v^*\|v_{i-1}^*]$	The expected effective value if agent A_i is about to start its exploration process after receiving a value v_{i-1}^*
$E[v^*]$	The expected effective value obtained eventually by each agent from the partnership
$P_i(j)$	The probability that agent A_i will eventually obtain, along its exploration process, the value of the opportunity associated with the j^{th} ($1 \leq j \leq n_i$) highest reservation value
$EC_i[cost\|v_{i-1}^*]$	The expected cost of agent A_i, given the value it receives v_{i-1}^*
$EC_i[cost]$	The expected accumulated cost of agent A_i
EB_i	The expected benefit of agent A_i