Utilizing Costly Coordination in Multi-Agent Joint Exploration[1]

Igor ROCHLIN [a,2], David SARNE [b]

[a] *Department of Computer Science*
Bar-Ilan University
Ramat-Gan, Israel
e-mail: igor.rochlin@gmail.com

[b] *Department of Computer Science*
Bar-Ilan University
Ramat-Gan, Israel
e-mail: david.sarne@gmail.com

Abstract. This paper studies distributed cooperative multi-agent exploration methods in settings where the exploration is costly and the overall performance measure is determined by the minimum performance achieved by any of the individual agents. Such an exploration setting can commonly be found in multi-agent systems, e.g., in multi-channel cooperative sensing where the quality of the overall connection is constrained by the individual qualities of the connections used by the different agents. The goal in such problems is to optimize the process as a whole, considering the tradeoffs between the quality of the solution obtained and the cost associated with the exploration and coordination between the agents. The methods considered in this paper differ in the level of coordination employed, ranging from no coordination to complete coordination. The strategy structure in all cases is shown to be threshold-based, and the thresholds which are analytically derived in this paper can be calculated offline, resulting in a very low online computational load. The analysis is extended to the case where coordination is supplied by a self-interested monopolist communication provider, charging a fee that depends on the number of agents for which coordination is required. In this case, the agents' expected-benefit-maximizing cooperative exploration strategy is to have some sub-groups coordinate their exploration (if at all) while the remaining agents explore individually with no coordination between them. We show that given the option for side-payments, the exploring agents can improve their expected benefit by compensating the communication provider to change the price at which she offers her services. An illustration for the importance of considering others' findings in one's strategy is given using the spectrum sensing application, experimenting with real data.

Keywords. Multi-Agent Exploration, Multilateral Search, Joint Exploration, Cooperation, Coordination, Dynamic Spectrum Access Networks

---

[1] Preliminary results of this work appeared in the Proceedings of the Twenty-Seventh Conference on Artificial Intelligence (AAAI-13).

[2] Corresponding Author

1. Introduction

In many settings, the benefits from the different alternatives available to an agent are associated with some uncertainty. For example, in eCommerce, a shopbot does not know a priori the pricing of a requested product among the different merchants' web-sites. Similarly, in the Mars exploration rover mission, the rover does not know a priori the terrain conditions in the different locations it can potentially visit. In both examples, the agents can explore the alternatives (denoted "opportunities" onwards) available to them, revealing the actual benefit ("value") with which they are associated, however incurring a cost (either monetary or in terms of consuming some of their resources) as such exploration is inherently costly.[3] 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 the opportunity eventually picked minus the costs accumulated along the exploration process. This kind of exploration is standard in autonomous-agent literatures [11,17,21].

This exploration process becomes more complex whenever conducted cooperatively by several agents. For example: when the agents are robots that need to evaluate several potential locations for mining a certain mineral on the face of Mars [17]; a group of buyers that need to evaluate several potential sellers for buying different products [37]; and secondary users in Dynamic Spectrum Access applications that need to evaluate different connections to a central server in order to establish a common communication link [1]. The cooperative exploration is more complex in the sense that the agents' exploration is now affected also by findings of other agents in the group. The agents thus need to coordinate their exploration. The key for coordination is the ability to communicate, and communication is inherently costly. Therefore the agents' cooperative exploration strategy also needs to take into consideration the overhead associated with the coordination between them along the process.

In this paper, we formally introduce and analyze a model of a cooperative multilateral exploration in which exploration and communication (and consequently coordination)[4] are inherently costly and the performance of each agent is affected by the results of the exploration carried out by all of the other agents. Specifically, we focus on a type of problems where the benefit of each agent is the worse (e.g., the minimum benefit or the maximum delay) among the best results obtained in any of the individual exploration efforts. For example, in Dynamic Spectrum Access applications each agent evaluates different channels through which it will connect to the central server that facilitates communication between the different agents; hence the quality of service experienced by all of the agents depends on the lowest-quality channel selected for the communication by any of the agents. Another example concerns settings where several individuals need to meet at a certain location on an ad-hoc basis. Upon deciding on the meeting, each individual may evaluate different alternatives for getting there as soon as possible,

[3]Our use of the term "costly" does not aim to impose it is overpriced but rather that it is "non-free".

[4]In the remaining of the paper we will use "coordination" and "communication" interchangeably.

however the earliest time the group will actually be able to meet is the longest duration it will take any of the individuals to get there. A similar setting arises 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 *minimum* function over the sub-tasks [42,3]. Overall, in cooperative exploration settings, the system's performance, which the agents attempt to maximize, is commonly the sum of the resulting individual benefits minus the costs (of exploration and coordination) accumulated by the different agents along the process. Therefore, given the cost of coordination, the agents may choose to have only some of them coordinate their exploration efforts, while others explore individually with no coordination whatsoever with the other group members.

The analysis provided in the paper is based on extracting a solution for two extreme cases: the first is when the cost of coordination is insignificant, thus all agents' exploration efforts are coordinated, and the second is when the cost is substantial to an extent that precludes any benefit from using it. These two cases are then integrated to a combined (hybrid) case that applies to any coordination cost structure, wherein not all the agents necessarily employ coordination. As discussed in details in the related work section, former cooperative exploration literature usually does not consider the overall quality to be a function of the individual qualities of the different agents or does not take into consideration the coordination costs. The very few models that do consider cooperative exploration settings with constrained quality functions constrain the exploration in a way that precludes the use of a hybrid exploration schemes of the type proposed in this paper.

The model is further extended for the case where the coordination (if any) that takes place between the agents is enabled through a third party (communication provider), who can self-interestedly set the fee for the coordination service. We show how the communication provider can set her expected-benefit-maximizing fee. One interesting implication of this analysis is that if an external side-payments mechanism can be used, then the transfer of payments from the agents to the communication provider, in a way that fully compensates her for setting a fee different from her expected-benefit-maximizing one, can result in an improvement of the agents' expected benefit.

In the following section we formally introduce the model. The model analysis is given in Section 3. Section 4 uses a tractable synthetic setting for illustrating the optimal (expected-benefit-maximizing) cooperative exploration strategy and its properties, as well as the effect of the different model parameters over the overall performance. An experiment with real data, in the area of spectrum access is reported in Section 5. Related work is reviewed in Section 6, emphasizing the uniqueness of the analysis provided in the paper. Discussion and conclusions are given in Section 7.

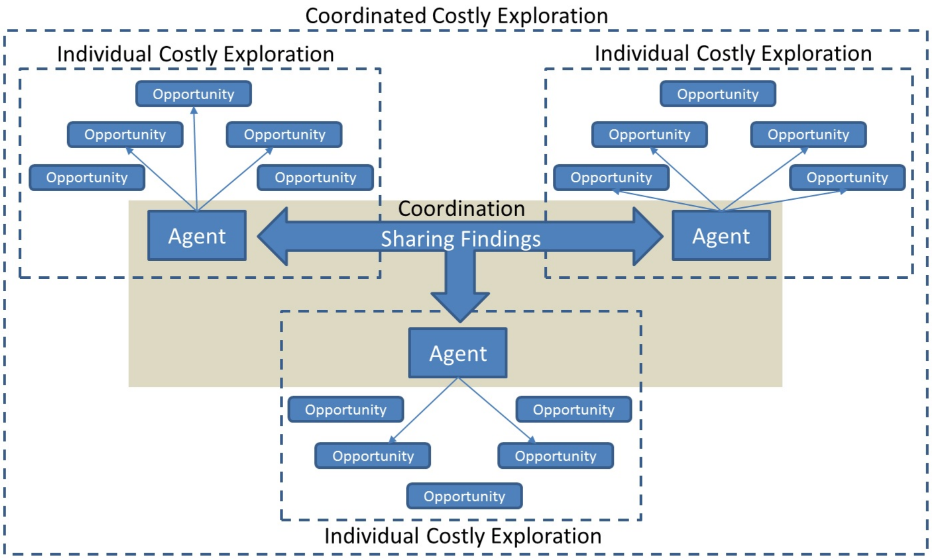

Figure 1. A schematic illustration of the cooperative coordinated costly exploration process - each agent executes an individual costly exploration and coordination is achieved through sharing findings.

2. The Model

Cooperative Costly Exploration. We consider a setting where a group of k cooperating agents attempt to achieve a shared goal.[5] In order for the goal to be achieved, each agent needs to engage in a *costly exploration* [11], i.e., to evaluate different opportunities which values are a priori uncertain (see Figure 1). This exploration of opportunities is considered costly (non-free) in the sense that eliminating the uncertainty associated with the value of any given opportunity incurs a cost c (expressed in terms of opportunity values), as the agent needs to consume some of its resources as part of this process. These individual exploration processes are standard [17,21,27]. The uncertainty associated with each opportunity's value is modeled, as in most costly exploration literature [8,10,26,30], through a probability distribution function $f(x)$, i.e., the value of each opportunity in any of the individual exploration processes is drawn from $f(x)$. The model assumes that the agents are not limited by the number of opportunities they can evaluate. Once all agents have completed their individual exploration, the benefit of each of the k agents from the resulting cooperative exploration process is the minimum among the best results obtained in any of the individual exploration processes, denoted v^*. The overall benefit is thus kv^* minus the costs accumulated along the individual explorations. Since the agents are cooperative, their goal is to maximize the overall expected benefit.

Taking the Dynamic Spectrum Access application domain as an example, each agent represents a terminal and all terminals are interested in establishing a connection between them (e.g. for a conference call, document/video sharing or a

[5] See Appendix 8 for a summary of all the notations used in this paper.

multi-player game). The terminals are located in different geographical locations and each terminal can use different wireless channels to connect to a central server supporting the requested application. Each terminal carries out a costly exploration process in the form of sensing different channels of different qualities until it selects a specific channel with a specific quality. The sensing is costly in that the terminal needs to allocate some of its resources for the task (e.g., energy or delay other transmissions). The quality of service provided by the application depends on the qualities of all individual channels selected (e.g., if one of the terminals ends up with a low quality channel, the experience of all of the users will be negatively affected). Hence, the quality of service provided to all of the terminals will be a function of the lowest quality channel selected by any of the terminals. As for the ad-hoc meeting application domain, here each agent represents an individual interested in having all group members arrive to the meeting place as early as possible, once the meeting place was determined (e.g., three hungry students decide spontaneously to get together and have lunch at a specific restaurant, hoping to get there as soon as possible). The agents are located in different geographical locations and each can choose between various different means of transportation for getting to the meeting place (e.g., by different train, subway or bus lines, by car, considering different alternative routes, or any combination of these means of transportation). Checking (i.e., exploring) an alternative potentially involves several activities (e.g., looking at the map, checking web-sites for route, checking timetables) thus is costly in the sense that it takes time (hence further delaying the arrival). The benefit from the findings depends on the time all agents have arrived to the meeting, hence it is the longest time it takes any of them getting there.

The agents are assumed to be fully rational and acquainted with the distribution function $f(x)$, the number of agents k and the exploration cost c. Their decision whether to evaluate an additional opportunity at any point along their exploration process thus needs to take into consideration the tradeoff between the marginal improvement in the value of v^* and the cost incurred along the process.

Costly Coordination. The model assumes the agents can coordinate their cooperative costly exploration through sharing their findings along the process (see Figure 1). The sharing of findings is typically facilitated through communication. Specifically, we assume that in order to share findings, the agents need to purchase coordination-facilitating communication services from an external operator (communication provider). The model assumes that the cost of supplying j agents with such coordination services between them (i.e., enabling j agents to share their findings) is given by the function $c_m(j)$. The function $c_m(j)$ is assumed to be non-decreasing in j. One natural cost function is $c_m(j) = c_m \cdot j$, meaning that the cost of supplying any additional agent with coordination capabilities throughout the exploration is c_m.[6] Given the cost of coordination, the goal of the agents is to decide on the set of agents that will be equipped with coordination capabilities

[6]This cost structure is typical whenever the coordination is executed in the form of a central server which distributes the information to the different subscribers. Alternatively, it can be justified in physical worlds by the cost of a specific HW, e.g., a push-to-talk handheld device, that will be used by any of the coordinating agents throughout the exploration.

(i.e., which exploration will be coordinated). For example, the agents can decide to have k' ($k' < k$) of them operate as one coordinated subgroup, k'' ($k' + k'' < k$) of them operate as a second coordinated subgroup and the remaining $k - k' - k''$ agents operate individually, each executing its exploration in isolation. The coordination cost in this case is $c_m(k') + c_m(k'')$. At the end of the process, the best results of all the k agents will be revealed and the individual agent benefit minimum among them. Therefore the agents' strategy in this case specifies both the subgroups of agents which search will be coordinated and the way each subgroup of agents will conduct its own coordinated exploration. The optimal strategy is the one resulting in the maximum expected overall benefit, defined as kv^* minus the coordination costs and the aggregate of costs accumulated along the individual explorations.

The communication provider is assumed to be self-interested and attempts to maximize her benefit, denoted B_p. The provider's benefit, if the agents request coordination services for w groups of sizes $k_1, .., k_w$ ($\sum k_i \leq k$), is given by: $B_p = \sum_{i=1}^{w}(c_m(k_i) - c_p(k_i))$, where $c_p(j)$ is the provider's cost of serving j exploring agents. The problem thus can be thought of as a Stackelberg game where the communication provider is the first mover, publishing its offered coordination service rates in the form of $c_m(j)$, and the agents are the followers.

3. Analysis

We first introduce the optimal (expected-benefit-maximizing) exploration strategy for a single agent facing the exploration problem, i.e., without restricting the value found by the other agents. We then augment that strategy and adapt it to the case of k cooperating agents with the minimum value restriction. Our analysis of the cooperative exploration considers three cooperative distributed exploration strategies. The first relates to the case where the coordination cost is substantially high, thus the agents prefer not to use the coordination service to any extent. The second is when the coordination is free (i.e., $c_m(j) = 0$ for every $j \leq k$), thus coordination is fully adopted. In both methods the opportunities are evaluated sequentially (in a random pre-defined order, as they are all a priori alike). The analysis of the two cases facilitates the analysis of the general case, for any coordination cost structure $c_m(j)$, wherein not all the agents necessarily employ coordination.

3.1. Optimal Costly Exploration Disregarding Others' Findings

When relaxing the minimum-value restriction, each agent's exploration problem can be analyzed separately and solved as a classic sequential exploration problem of the kind widely used in search theory [20,19,27,15,47]. In sequential exploration, a single agent faces a stream of opportunities that arise sequentially, incurring a cost c for revealing the value of each, where the values are associated with a probability distribution function $f(y)$. The agent needs to set her strategy as the mapping from the value received to the set $\{terminate, resume\}$, i.e., a stopping rule. The optimal stopping rule in this case, is a reservation-value rule [47,26] (i.e.,

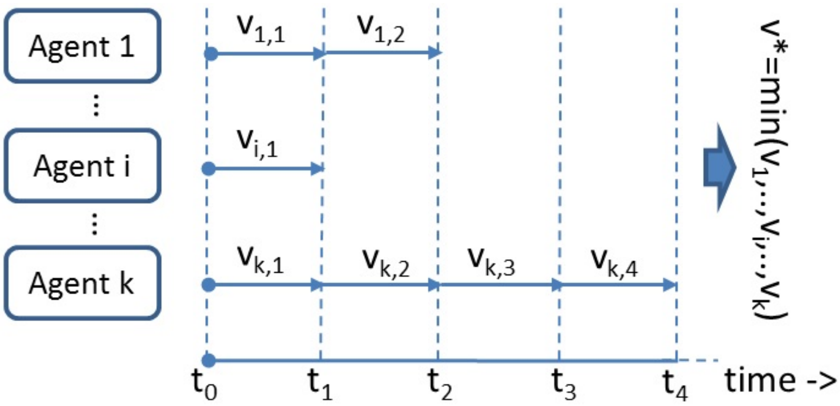

Figure 2. A schematic illustration of the strategy that does not use coordination, where $v_{i,j}$ denotes the value obtained by agent i at time j, if the agent was exploring an opportunity then. Here, $v_1 = \max(v_{1,1}, v_{1,2})$, $v_i = v_{i,1}$ and $v_k = \max(v_{k,1}, v_{k,2}, v_{k,3}, v_{k,4})$.

a threshold): the agent terminates the exploration once a value greater than a reservation value r is revealed, where the expected-benefit-maximizing reservation value r satisfies:

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

Intuitively, r is the value where the agent is precisely indifferent: the expected marginal benefit from obtaining the value of the opportunity exactly equals the cost of obtaining that additional value. Interestingly, the decision rule expressed by the optimal strategy according to (1) is myopic, i.e., the value of r does not depend on the number of opportunities that can still be potentially explored [47] but rather only on the characteristics of the opportunity (distribution of values and exploration cost).

We denote the above exploration strategy as "disregarding others' findings" in the context of the multi-agent exploration with value restriction, since it does not take into consideration the exploration of the other agents and the resulting influence over the effective value. 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 one given above, i.e, carried out sequentially according to reservation values, though the reservation values used are different.

3.2. Cooperative Non-Coordinated Costly Exploration

If the agents cannot coordinate whatsoever, then their exploration takes place separately and can be perceived as performed in parallel, where the value of v^* is revealed only after all individual explorations have come to an end (see Figure 2). The optimal cooperative strategy in this case is based on having each agent use a reservation value for its individual exploration. Each agent will keep exploring as long as the best value found is below the reservation value it uses. The optimality (in terms of overall expected-benefit-maximization) of the reservation-value-based strategy in this case derives from the fact that the agent's state depends solely on the best value found so far. Therefore, if the agent finds it beneficial to explore when the best value obtained so far is v, then it is inevitably beneficial to explore if the best value found so far is v' for any $v' < v$ (and vice versa).[7]

Theorem 1. *The cooperative optimal exploration strategy with no coordination between the agents is to have each agent A_i use a reservation value r that satisfies:*

$$c = k\int_{y=r}^{\infty} f(y)\Big(\int_{x=-\infty}^{\infty}(min(y,x) - min(r,x))\bar{f}(x)dx\Big)dy \tag{2}$$

where $\bar{f}(x)$ is the probability distribution function of the minimum among the best values obtained by all other agents, i.e, $min(x_1,..,x_{i-1},x_{i+1},..,x_k)$.

Proof. The overall expected benefit of the system from the cooperative exploration is the expected value with which the agents end up, denoted EV, multiplied by k, minus the accumulated costs along the exploration. Now consider the effect of agent A_i's exploration over EV. Given the distribution of the minimum among the best values obtained by all other agents, $\bar{f}(x)$, the value of EV when A_i uses a reservation value r is given by:

$$EV = \int_{y=-\infty}^{r} EV f(y)dy + \int_{y=r}^{\infty} f(y)\int_{x=-\infty}^{\infty} min(y,x)\bar{f}(x)dxdy \tag{3}$$

The above recursive equation captures the value of EV when agent A_i evaluates an additional opportunity (i.e., when the best value the agent obtained so far is below r) hence obtaining an additional value y. The first term relates to the case where a value $y < r$ is obtained through this additional exploration, in which case agent A_i resumes its exploration process according to the reservation value rule. Since A_i is not limited by the number of opportunities it can evaluate, it now faces the exact same decision problem as before, resulting in an expected value EV. The second term relates to the case where the value y obtained is above r, in which case the agent terminates its exploration and the expected value EV is the minimum value found among the value obtained by A_i, i.e., y, and the minimum best value x obtained by the other agents (captured by the distribution function $\bar{f}(x)$).

The expected number of opportunities explored by agent A_i is $\frac{1}{1-F(r)}$ as this becomes a Bernoulli sampling process with a success probability of the value obtained being greater than the threshold used, i.e., $1 - F(r)$. Consequently, the expected cost accumulated along the exploration of A_i when using r is given by $\frac{c}{1-F(r)}$. Therefore, the overall expected benefit of the system as a function of the

[7]The myopic and stationary natures of the reservation value derive from the fact that since neither agent is limited by the number of opportunities it can explore, the rejection of the current opportunity results in facing the exact same decision problem once again in the next exploration round.

reservation value r used by A_i, denoted $B(r)$, is given by (after isolating EV in Equation 3) and multiplying it by k:

$$B(r)=\frac{k\int_{y=r}^{\infty}f(y)\int_{x=-\infty}^{\infty}min(y,x)\bar{f}(x)dxdy}{1-F(r)}-\frac{c}{1-F(r)}-C \tag{4}$$

where C denotes the expected cost accumulated along the other agents' exploration. In order to find the optimal reservation value r, we take the first derivative of Equation 4 and set to zero:

$$\frac{dB(r)}{dr} = \frac{-k(f(r)\int_{x=-\infty}^{\infty}min(r,x)\bar{f}(x)dx)(1-F(r))}{(1-F(r))^2} \tag{5}$$
$$+\frac{kf(r)\int_{y=r}^{\infty}f(y)\int_{x=-\infty}^{\infty}min(y,x)\bar{f}(x)dxdy}{(1-F(r))^2}-\frac{cf(r)}{(1-F(r))^2}=0$$

which after some mathematical manipulations becomes:

$$c = k\int_{y=r}^{\infty}f(y)\Big(\int_{x=-\infty}^{\infty}min(y,x)\bar{f}(x)dx\Big)dy - k(1-F(r))\int_{x=-\infty}^{\infty}min(r,x)\bar{f}(x)dx \tag{6}$$

and since $(1-F(r))=\int_{y=r}^{\infty}f(y)dy$, Equation 6 is equivalent to (2). $\square$

Equation 2 also has the intuitive interpretation, as in 3.1, in the form of indifference between the expected marginal utility from obtaining the value of the opportunity (represented by the right-hand term of the equation), this time, however, calculated for all group members (hence multiplied by k) and taking into consideration the findings of others.

Since all agents face a similar setting (characterized by $f(x)$ and c) they all use the same reservation value r. This enables a simple formulation of the function $\bar{f}(x)$:

$$\bar{f}(x) = \frac{d(1-(1-F^{returned}(x))^{k-1})}{dx}, \tag{7}$$

where:

$$F^{returned}(x) = \begin{cases} 0 & x \le r \\ \frac{F(x)-F(r)}{1-F(r)} & x > r \end{cases} \tag{8}$$

The function $F^{returned}(x)$ returns the probability that the best value with which an agent that uses a reservation value r ends up (i.e., in its individual exploration) will be below x. The term $(1-(1-F^{returned}(x))^{k-1})$ is thus the probability that the minimum among the results of the other $k-1$ agents' explorations will turn out to be below x, and therefore its derivative is the probability distribution function of the minimum among the best values obtained by all other agents.

Using Equation 2 we can now calculate r, and since all agents use the same reservation value, the probability distribution function of the minimum among

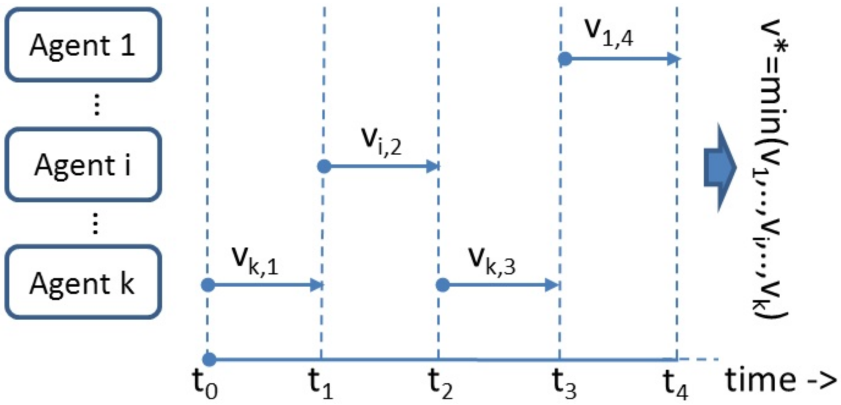

Figure 3. A schematic illustration of the fully coordinated exploration strategy, where $v_{i,j}$ denotes the value obtained by agent i at time j, if the agent was exploring an opportunity then. Here, $v_1 = v_{1,4}$, $v_i = v_{i,2}$ and $v_k = \max(v_{k,1}, v_{k,3})$.

the best values found by *all* agents (unlike with $\bar{f}(x)$ which apply to all agents except one) is given by: $\frac{d(1-(1-F^{returned}(x))^k)}{dx}$. Therefore, the expected value the agents end up with, EV, is thus given by:

$$EV = \int_{x=-\infty}^{\infty} \big(x \cdot \frac{d(1-(1-F^{returned}(x))^k)}{dx}\big)dx \tag{9}$$

As shown in the proof for Theorem 1, the accumulated cost along the exploration process of each of the agents is given by $\frac{c}{1-F(r)}$, hence the system's overall expected benefit, denoted EB, is:

$$EB = k\big(EV - \frac{c}{1-F(r)}\big) \tag{10}$$

3.3. Cooperative Fully Coordinated Costly Exploration

If the agents can coordinate without incurring a cost, then their exploration strategy should take into consideration, at each step of the process, the values found by any of the other agents. Furthermore, since the exploration is costly, it is advantageous for the agents to execute their exploration sequentially (having one agent explore at a time) rather than in parallel (see Figure 3). Since the state of the agents now depends on the vector of best values found by the different agents, the optimal strategy is no longer based on a single reservation-value. Instead, as we prove in this section, it assigns a different reservation value for each state and applies it to (i.e., compares it with) the minimum among the set of best values found by the different agents.

We represent the system's state as a vector $V = (v_1, .., v_k)$, in the k-dimensional space, where v_i ($1 \leq i \leq k$) is the best value found so far by agent A_i.[8] We use $S(V) \rightarrow \{i, terminate\}$ to denote the agents' strategy, where i

[8] If agent A_i has not yet engaged in an exploration then $v_i = 0$.

$(1 \leq i \leq k)$ suggests that agent A_i needs to execute an exploration step next and *terminate* means the exploration as a whole should be terminated. If the exploration is terminated, then the value of v^* is determined according to the minimum value of V. For convenience we use V_{min} to denote the minimum value in V (i.e., $V_{min} = min(v_1, .., v_k)$). Due to the nature of v^* it is obvious that given a state V, if the optimal strategy is to resume the exploration, then the agent who should be evaluating an additional opportunity is the one whose highest value is V_{min}.[9] This is simply because any increase in the best value obtained by any other agent A_i associated with a best value $v_i > V_{min}$ can affect v^* only if it is accompanied by findings greater than v_i of agents currently associated with best values lower than v_i. The agents' strategy can therefore be expressed as: $S(V) \rightarrow \{resume, terminate\}$.

Proposition 1. *For any state $V = (v_1, .., v_k)$, if $S(V) = resume$ then inevitably $S(V') = resume$ for any V' differing from V only in the value of its minimum element, where $V'_{min} < V_{min}$ (formally: $V' = (v_1, ..., v_{i-1}, v'_i < v_i, v_{i+1}, ..., v_k)$ where $v_i = V_{min}$).*

Proof. Assume otherwise, i.e., $S(V') = terminate$. Since $S(V) = resume$, then once a state V is reached, the agent associated with V_{min} will resume exploring until a better value is obtained. From the system's point of view, this is preferable over terminating the exploration, i.e., over ending up with a value V_{min}. Now consider the option to resume exploration by the agent associated with value V'_{min} when starting from a state V' until obtaining a value greater than V_{min}. The expected exploration cost and the distribution of the value with which the agent will end up (i.e., above V_{min}) is equal in both cases. Therefore, since terminating the exploration process when in state V' yields $V'_{min} < V_{min}$, the strategy $S(V') = terminate$ cannot be optimal (in terms of expected-benefit-maximization). □

The main implication from Proposition 1 is that for all states that differ only in the value of their minimum element there is a single reservation value for determining whether to resume exploration. Whenever reaching a new state, each agent needs to determine if the best value it had obtained so far is V_{min}. The agent associated with V_{min} will calculate the reservation value according to the current state V, denoted $r(V)$, and resume the exploration if its value is below $r(V)$.

We use $\sigma(V, y) \rightarrow V'$ to denote the new state to which the system transitions, if it was initially in state V, after the agent associated with V_{min} has obtained a value y in its exploration. If v_i is the minimum value in V (i.e., $v_i = V_{min}$) then:

$$\sigma(V, y) = \begin{cases} V & y \leq V_{min} \\ (v_1, ..., v_{i-1}, y, v_{i+1}, ..., v_k) & \text{otherwise} \end{cases} \tag{11}$$

The system's expected benefit onwards, when in state V, denoted $EB(V)$, if continuing according to the optimal strategy $S(V)$ is thus given by the following recursive equation:

[9] In the case where two or more agents hold a value equal to V_{min} there is no importance to the selection of which of them will explore next.

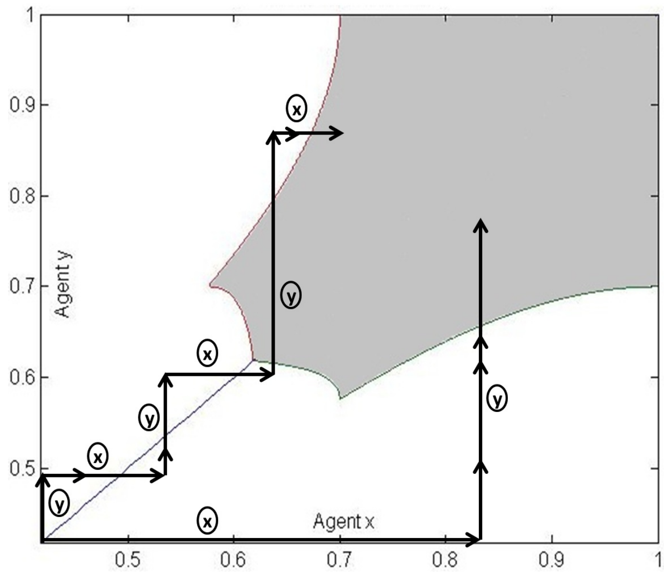

Figure 4. An example of 2-agent cooperative exploration with full coordination. The agents are A_x and A_y, the exploration cost is $c = 0.09$ and the distribution of values is uniform ($f(x) = 1$, over the interval $(0, 1)$ and zero otherwise). The arrows represent two possible exploration paths, where the letter in the circle is the exploring agent. The gray area represents all states in which the exploration terminates.

$$EB(V)=\begin{cases} kV_{min} & r(V)\leq V_{min} \\ -c+\int_{y=\infty}^{r(V)} EB(V)f(y)dy + \int_{y=r(V)}^{\infty} EB(\sigma(V,y))f(y)dy & \text{otherwise} \end{cases} \tag{12}$$

where in the case $r(V)\leq V_{min}$ the exploration necessarily terminates according to the reservation value rule, hence the benefit is kV_{min}. For other cases, the exploration resumes, thus incurring a cost c and obtaining a new value y. The new state is thus either V (if $y \leq r(V)$) or $\sigma(V, y)$ (if $y > r(V)$).

Similarly we can formulate the expected overall number of opportunities explored, when in state V, denoted $EN(V)$:

$$EN(V)=\begin{cases} 0 & r(V)\leq V_{min} \\ 1+\int_{y=\infty}^{r(V)} EN(V)f(y)dy + \int_{y=r(V)}^{\infty} EN(\sigma(V,y))f(y)dy & \text{otherwise} \end{cases} \tag{13}$$

and the expected overall number of opportunities explored is given by $EN(0, ..., 0)$.

Theorem 2. *The optimal exploration strategy with full coordination, when in state V, is to terminate the exploration if $V_{min}\geq r(V)$ and otherwise resume exploration, where $r(V)$ is the solution to:*

$$c = \int_{y=r(V)}^{\infty} (EB(\sigma(V, y)) - kr(V))f(y)dy \tag{14}$$

The value $r(V)$ is the same for all states differing only by their minimum value (thus can be calculated only once for these states).

Proof. In order to find the optimal (in terms of overall expected-benefit-maximization) reservation value $r(V)$ we take the first derivative of Equation 12 and set it to zero:

$$\frac{dEB(V)}{dr(V)} = f(r(V))EB(V) - f(r(V))EB(\sigma(V, r(V))) = 0$$

resulting in $EB(V) = EB(\sigma(V, r(V)))$. Since $EB(\sigma(V, r(V))) = kr(V)$ according to (12), and substituting $EB(V)$ according to (12), we obtain:

$$-c + kr(V)F(r(V)) + \int_{y=r(V)}^{\infty} EB(\sigma(V, y))f(y)dy = kr(V) \tag{15}$$

and since $(1 - F(r(V))) = \int_{y=r(V)}^{\infty} f(y)dy$, Equation 15 is equivalent to (14). $\square$

The main implication of Theorem 2 is that the agents do not need to assign a different reservation value to each vector of best values found. Instead only the reservation value of states differing by their minimum value needs to be calculated. The "reservation" property in this case is thus actually of the form of a reservation "frontier", i.e., a sub-space of the state space where the exploration should be terminated. Figure 4 depicts the strategy space in a two-agent cooperative exploration with full coordination. Each point in the two-dimensional space corresponds to a different state, differing in the best value obtained so far by Agent X (horizontal axis) and Agent Y (vertical axis). The gray and white areas represent states for which the optimal strategy is to terminate and to resume exploration, respectively. For any state in the white area below the diagonal, the agent that will perform the exploration is Agent Y (and above the diagonal, Agent X).

The overall expected benefit of the system, EB, can be calculated using Equation 12 when starting from the state where none of the agents has engaged in exploration yet, i.e., $EB = EB(0, .., 0)$.

3.4. Cooperative Coordinated Costly Exploration with Costly Coordination

Finally, we analyze the general case, i.e., where $c_m(j)$ is a general non-decreasing function, wherein not all of the agents necessarily employ coordination. In this case, it is possible that only a subgroup of agents will use coordination or that the agents will divide into sub groups that coordinate their exploration separately. For some cases, e.g., when $c_m(j)$ is linear in j or increases in a decreasing rate as j increases, the optimal strategy is to have at most one group of agents coordinate their exploration and to have the remaining agents execute their exploration in isolation. This is because whenever two groups of agents merge and coordinate their exploration, the overall coordination cost does not increase and the same performance as with the two separate groups can be achieved even by using a sub-optimal solution according to which each agent follows the exploration strategy it would have used if operating in its original group (hence the division into groups cannot possibly improve the performance of a single unified group). In other (though less common) cases, the optimal strategy is to have several subgroups of agents coordinating their exploration separately (and potentially also some individuals executing exploration completely by their own).

Consider the case where the k agents request coordination services for w subgroups of sizes $\{k_1, ..., k_w\}$ ($\sum k_i \leq k$), such that the exploration of the agents within each subgroup is coordinated and each of the remaining $k - \sum k_i$ agents explore in isolation. We use $r_i(V)$ to denote the reservation value function used by the agents of the i-th subgroup, upon reaching a state $V = \{v_1, ..., v_{k_i}\}$ in its coordinated exploration. We use $F_i^c(V, x)$ to denote the probability that the agents in the i-th subgroup, exploring with coordination, end up with a minimum value of x or below, given that they start from state V. The function $F_i^c(V, x)$ can be calculated recursively according to:

$$F_i^c(V,x) = \begin{cases} 0 & x < V_{min} \\ 1 & x \geq V_{min} \wedge r_i(V) \leq V_{min} \\ \int_{y=-\infty}^{\infty} F_i^c(\sigma(V,y),x) f(y) dy & \text{otherwise} \end{cases}$$

The case where $x < V_{min}$ is trivial since given a state V, the minimum value the agents end up with is at least V_{min}. Therefore, if $x < V_{min}$ the probability of ending up with x or below is zero. Similarly, when $r_i(V) \leq V_{min}$ the agents in the i-th subgroup necessarily terminate and their value v^* is V_{min}. Hence, since $V_{min} \leq x$ the function obtains 1. In all other cases, the exploration resumes, hence the probability is given recursively based on the new state $\sigma(V, y)$ the agents will be in after the following exploration.

Using $F_i^c(V, x)$ we can calculate the probability that the agents in the i-th subgroup end up with a minimum value of x or below, denoted $F_i^c(x)$, as: $F_i^c(x) = F_i^c((0, ..., 0), x)$. The latter function enables calculating the probability distribution function of the minimum of the best findings of all agents other than those in the i-th subgroup, denoted $\bar{f}_i^c(x)$. This probability distribution function is, by definition, the first derivative of the probability that the minimum among the values returned by all agents other than those in the i-th subgroup is equal to or lower than x, i.e.: $\bar{f}_i^c(x) = \frac{d(1-(1-F^{returned}(x))^{k-\sum k_i} \prod_{j \neq i}(1-F_i^c(x)))}{dx}$, where $F^{returned}(x)$ is calculated according to (8). Finally, we can calculate the probability distribution function of the minimum of the best findings of all agents, except for one single agent out of the $k - \sum k_i$ agents that explore in isolation, denoted $\bar{f}^c(x)$, given by: $\bar{f}^c(x) = \frac{d(1-(1-F^{returned}(x))^{k-\sum k_i - 1} \prod(1-F_j^c(x)))}{dx}$.

Based on the above probability distribution functions, we can now calculate the expected minimum value the k agents end up with when the process terminates, EV. For this, we first calculate, $EV_i(V)$, the expected minimum value with which the agents end up given that the agents of the i-th subgroup are in state V and no other a priori information regarding the findings of the other agents. This value can be calculated as a modification of Equation 12, however taking into consideration the results of the agents belonging to all other subgroups, thus is given by:

$$EV_i(V)=\begin{cases}\int_{y=-\infty}^{r_i(V)}EV_i(V)f(y)dy+\int_{y=r_i(V)}^{\infty}EV_i(\sigma(V,y))f(y)dy & r_i(V)>V_{min}\\ \int_{x=-\infty}^{\infty}min(x,V_{min})\bar{f}_i^c(x)dx & \text{otherwise}\end{cases} \quad (16)$$

The first term in (16) in case $r_i(V) > V_{min}$ relates to the case where a value $y < r_i(V)$ is obtained through an additional exploration, in which case the agents of the i-th subgroup resume their exploration process according to the reservation value rule. In this case, the expected value onwards is $EV_i(V)$, as the state remains V. The second term relates to the case where the value obtained is above $r_i(V)$, in which case the agents of the i-th subgroup resume their exploration process according to the new state $\sigma(V, y)$. In this case, the expected value is $EV_i(\sigma(V, y))$. For the case $r_i(V) \leq V_{min}$ the agents of the i-th subgroup terminate their exploration and the expected value is the minimum between the value V_{min} obtained by the agents of the i-th subgroup itself and the expected minimum value returned by the other $k - k_i$ agents (associated with a distribution $\bar{f}_i^c(x)$). The value of EV can thus be calculated as $EV = EV_i(0, .., 0)$ (resulting in the same value for any i).

Now consider the i-th coordinated subgroup when reaching a state V. The expected overall cost onwards, denoted $EC_i(V)$, of that subgroup if using a reservation value function $r_i(V)$, can be calculated using a recursive equation similar to (16):

$$EC_i(V)=\begin{cases}-c+\int_{y=\infty}^{r_i(V)}EC_i(V)f(y)dy+\int_{y=r_i(V)}^{\infty}EC_i(\sigma(V,y))f(y)dy & r_i(V)>V_{min}\\ 0 & \text{otherwise}\end{cases} \quad (17)$$

The case $r_i(V) > V_{min}$ is a simple modification of (16). For the case $r_i(V) \leq V_{min}$, no further exploration needs take place. As for the $k - \sum k_i$ agents that explore with no coordination, the expected cost of each of them is $\frac{c}{1-F(r)}$. Therefore, the expected overall exploration cost, denoted EC, is given by: $EC = \sum_{i=1} EC_i(0, ..., 0) + (k - \sum k_i) \cdot \frac{c}{1-F(r)}$.

The overall expected benefit of the system, EB, can be calculated using Equations 16 and 17:

$$EB = kEV - EC - \sum_{i=1}^{w} c_m(k_i) \quad (18)$$

At this point, we have everything needed to introduce Theorem 3, which specifies the optimal exploration strategy for the case where the coordination is not completely free.

Theorem 3. *The optimal exploration strategy when the agents use coordination for w subgroups of sizes $\{k_1, ..., k_w\}$ ($\sum k_i \leq k$), is to set a reservation value r for the $k - \sum k_i$ agents exploring separately and a reservation value function $r_i(V)$ for the k_i agents of the i-th subgroup, for any $i \leq w$, according to the solution of the set of equations:*

$$c = k\int_{y=r_i(V)}^{\infty}(EV_i(\sigma(V,y)) - EV_i(r_i(V)))f(y)dy \tag{19}$$

(one equation for each subgroup i ($i \leq w$)), and:

$$c = k\int_{y=r}^{\infty}f(y)\Big(\int_{x=-\infty}^{\infty}(min(y,x) - min(r,x))\bar{f}^c(x)dx\Big)dy \tag{20}$$

Proof. The proof relies in large on the proofs given for the two previous cases (Theorems 1 and 2), therefore we only detail the differences. Equation 19 augments Equation 14 in a way that considers the effect of the minimum best value found by any of the agents that explore separately in parallel (i.e., without coordination). It is obtained by taking the first derivative of Equation 18 according to $r_i(V)$, equating it to zero and applying some standard mathematical manipulations. Equation 20 augments Equation 2 in a way that takes into consideration in $\bar{f}^c(x)$ the minimum value found by the agents exploring in coordination in addition to the minimum among those exploring individually in parallel. □

In order to find the optimal exploration strategy, we need to extract the overall expected benefit for potentially all possible divisions of the agents into disjoint (non-overlapping) subgroups. The division for which the highest expected benefit, EB, is obtained is the one by which the agents should operate. As discussed at the beginning of the section, for many $c_m(j)$ cost functions (e.g., for the linear case which is highly common), the optimal division of agents is necessarily of the form of one coordinated group of size $k' \leq k$, and the remaining agents exploring in isolation. Therefore in this case only k possible solutions need to be evaluated (i.e., for k' =0,2,3,..., k) according to Theorem 3. However, even when the cost function $c_m(j)$ is of a different form, the exponential (combinatorial) number of divisions that need to be evaluated is of small concern, since for most real-life applications the number of exploring agents is relatively moderate.

3.5. Communication Provider's Considerations

The goal of a self-interested communication provider is to maximize her expected benefit, defined in Section 2 as $B_p = \sum_{i=1}^{w}(c_m(k_i) - c_p(k_i))$, where $k_1, .., k_w$ are the sizes of the w groups for which coordination is requested and $c_p(j)$ is the provider's cost of serving j exploring agents, through the determination of the service fee $c_m(j)$. The change in $c(j)$ results in a different number and sizes of agent subgroups for which the service is requested (and purchased). While this may imply that the communication provider needs to optimize over all possible $c_m(j)$ non decreasing functions, for many functions the optimization is substantially simpler. For example, for the common case where $c_m(j)$ is linear in j (meaning that each additional coordinated agent within the group increases the coordination fee by a fixed amount c_m) this can be done in $O(k)$. This is because, as discussed at the beginning of Section 3.4, the agents in this case are divided into one group of coordinated agents, of size k', and the remaining agents search individually in isolation from all others. The number of agents using coordination as a function of c_m in this case is a step function and an increase in c_m can only result in a decrease in the size of the group for which coordination is sought. Therefore, the provider's expected benefit as a function of c_m can be divided into k segments, each relating to a different k' value, where B_p monotonically increases within each segment, with a sharp change (a "step") in the transition between segments (this is illustrated for a specific case in Figure 8 in the next section). Therefore the provider only needs to consider the set (of size k) of c_m values for which the agents transition to a new k' value. The value c'_m for which the agents will switch from k' to $k'-1$ equals the difference in EB when calculated with $c_m = 0$ for k' and $k'-1$ according to the guidelines given in Section 3.4. This is because this latter value captures the agents' surplus from having another agent coordination (in addition to the $k'-1$ already coordinated).

In such settings, if an external side-payments mechanism can be used, the agents may find it beneficial to request the provider to reduce the fee she charges for the coordination service from its benefit-maximizing fee structure, $c^*_m(j)$, to a socially optimal fee structure $c^s_m(j)$. In exchange, the agents will fully compensate the provider for the difference $\sum_{i=1}^{w^*}(c^*_m(k^*_i) - c_p(k^*_i)) - \sum_{i=1}^{w^s}(c^s_m(k^s_i) - c_p(k^s_i))$, where $k^*_1, .., k^*_w$ are the sizes of the w^* groups for which coordination is requested with the fee structure $c^*_m(j)$ and $k^s_1, .., k^s_w$ are the sizes of the w^s groups for which coordination is requested with the fee structure $c^s_m(j)$. The fee structure the agents will request to switch to is necessarily $c_p(j)$. This is because the agents and the provider actually share the benefit produced through the cooperative exploration, after subtracting the actual coordination expense $\sum_{i=1}^{w} c_p(k_i)$. The maximum "net" benefit is produced if the agents make their decision regarding the extent of coordination based on the "true" cost of coordination $c_p(j)$. Therefore, since the agents only need to ensure that the provider's benefit remains the same, an increase in the overall net benefit will necessarily increase their expected benefit.

4. Numerical Illustration

In order to illustrate the performance achieved with the different methods, we use a tractable synthetic setting that simplifies calculations yet enables demonstrating the main solution characteristics in a clean manner (i.e., eliminating the need to isolate external phenomena that are commonly present in simulated or real-life applications). The setting uses a uniform distribution function defined over the interval $(0, 1)$ (i.e., $f(x) = 1$, $\forall 0 \leq x \leq 1$ and zero otherwise). The coordination fee set by the communication provider is taken to be linear, i.e., $c_m(j) = j \cdot c_m$.

We first demonstrate the improvement achieved by taking into considerations other agents exploration in one's strategy. For this purpose we introduce Figure 5 which depicts the expected benefit (per-agent) (graph (a)) and reservation value (graph (b)) as a function of the number of agents, in a setting $c = 0.2$, for the case where agents disregard others' findings (which solution principles, as given in costly exploration literature, are summarized in 3.1) and the cooperative non-coordinated method (analyzed in 3.2). We emphasize that the latter method is the one offering the least performance improvement, as when the coordination cost is zero, as in this example, with some or full coordination better improvement can be achieved. From the figure we observe that the performance improvement, even

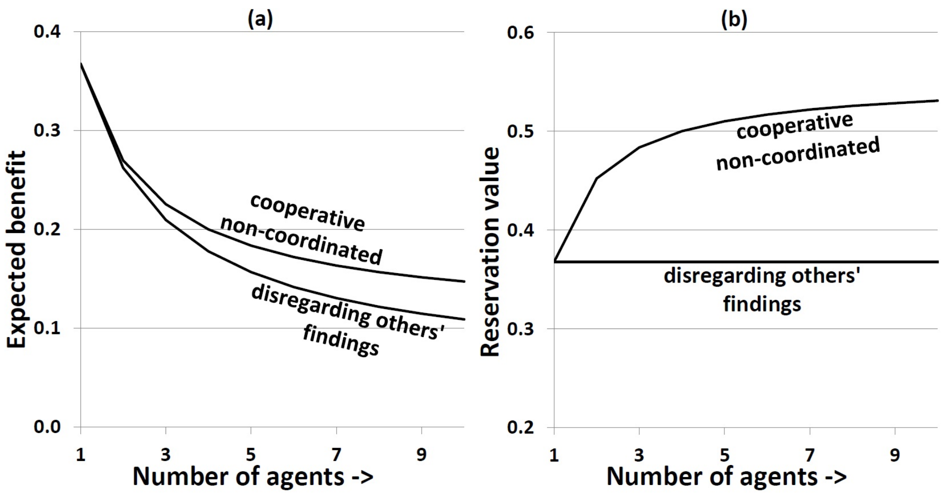

Figure 5. (a) Expected benefit as a function of the number of agents. (b) Reservation value as a function of the number of agents. Both settings use $c = 0.2$ and $c_m = 0$.

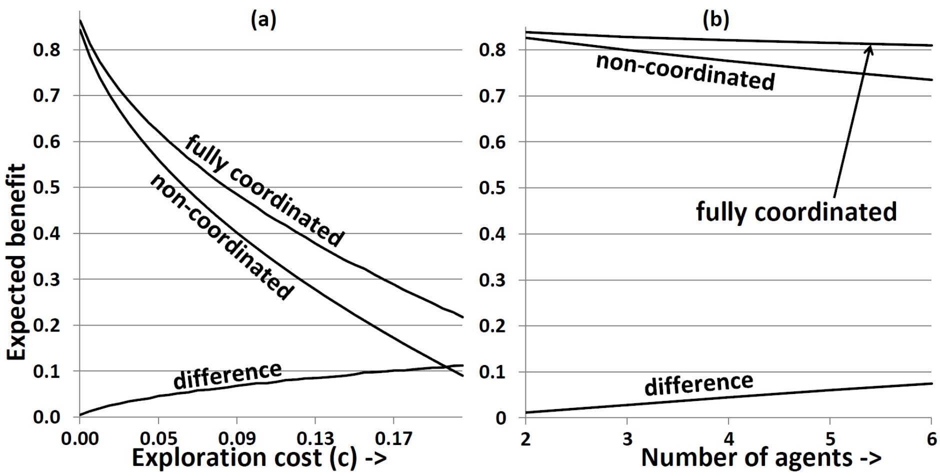

Figure 6. (a) Expected benefit (per agent) in a 3-agents' exploration as a function of the exploration cost. (b) Expected benefit (per agent) in a k agents' exploration as a function of k (using $c = 0.01$). Both settings use $c_m = 0$.

with the non-coordinated method, can be substantial. As expected, the greater the number of agents involved, the greater the reservation value used by the agents when taking into consideration others' findings. This is explained by the fact that as the number of agents increases, the expected minimum among the best values obtained individually decreases, therefore there is a greater benefit in increasing the reservation value used by the agents.

Figure 6 depicts the effect of the increase in the exploration cost c and the increase in the number of agents exploring cooperatively over the individual (per agent) expected benefit using $c_m = 0$. As expected, the fully coordinated case

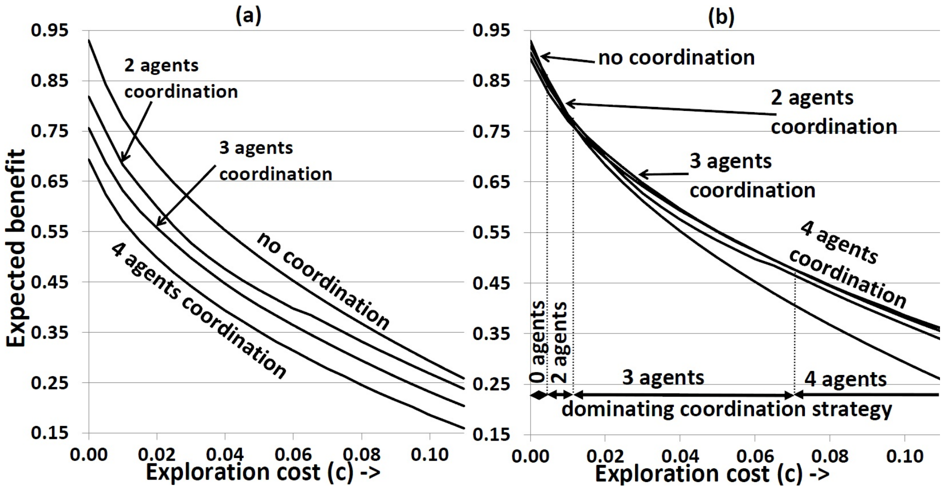

Figure 7. Expected benefit (per agent) as a function of the exploration cost, for $k = 4$, where the cost of coordination services is: (a) $c_m = 0.25$; (b) $c_m = 0.1$.

dominates exploration with no coordination, as far as expected benefit is concerned, and both the increase in exploration costs and in the number of agents exploring cooperatively result in a decrease in the expected benefit (per agent). The correlation between the expected benefit and the number of agents is explained by the fact that as the number of agents increases, the expected minimum of the obtained values decreases and more exploration is required. The curve marked "difference" depicts the difference between the individual expected benefit when the agents are using full coordination and when exploring with no coordination at all. When multiplied by the number of agents, this is the usage fee the communication provider should charge for an unlimited usage of coordination (i.e., for k agents' coordination), if its goal is to maximize expected revenue. This is also an upper bound for the value of coordination in this case.

Figure 7 depicts the expected benefit as a function of the exploration cost for $k = 4$, when each curve depicts a different number of agents using coordination services. The fee for coordination services is $c_m = 0.25$ (Figure 7(a)) and $c_m = 0.1$ (Figure 7(b)), and the agents can choose to have $0, 2, 3, 4$ of them operate in coordination. As expected, when the coordination cost is high, agents will explore in parallel with no coordination (Figure 7(a)), and when low, coordination is preferred to different extents (Figure 7(b); the number of agents using coordination is depicted at the bottom). From Figure 7(b) it is notable that the choice of how many agents will use coordination depends also on the exploration cost — when the exploration cost is low, agents can compensate over the lack of costly coordination with cheap extended exploration. As the exploration cost increases, the value of coordination increases as the saving achieved in repeated exploration becomes substantial.

Finally, Figure 8 illustrates the benefit of side-payments, using a setting where: $k = 4$, $c = 0.1$ and $c_p = 0.08$. Figure 8(a) depicts the communication provider's benefit as a function of the coordination cost c_m she uses. The benefit is determined according to the number of agents that uses the coordination

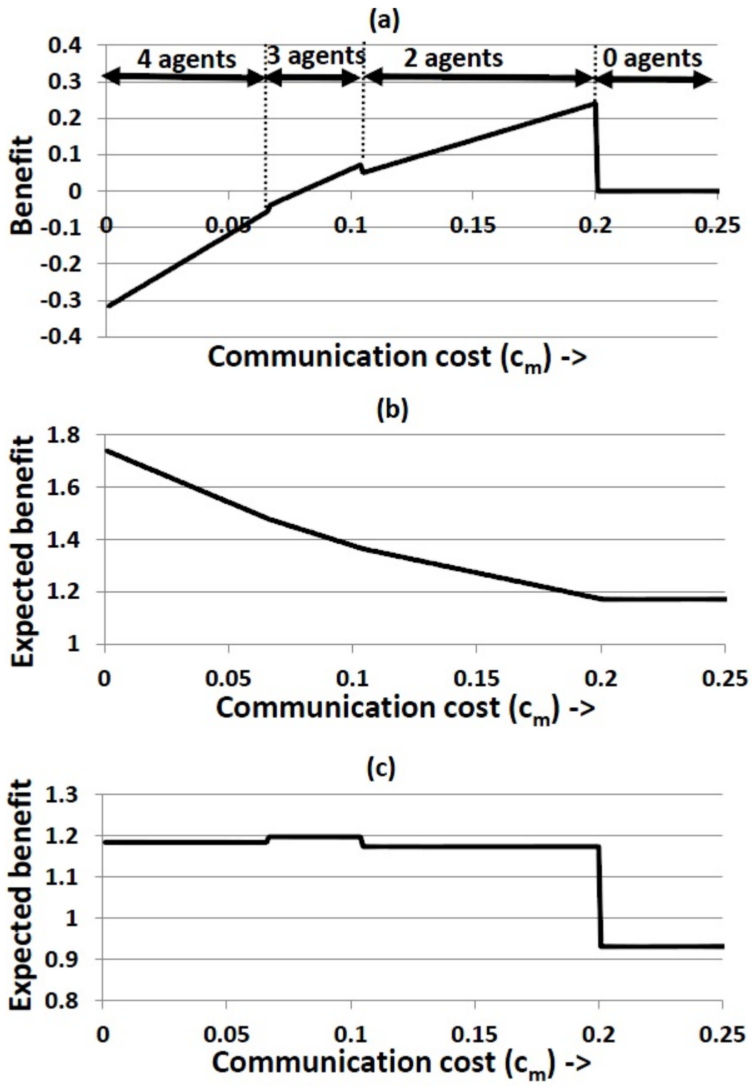

Figure 8. (a) Provider's benefit as a function of the coordination cost c_m set. (b) Agents' overall expected benefit as a function of the coordination cost c_m set. (c) Agents' overall expected benefit when side-payments are used, as a function of the coordination cost c_m set. The setting used is: $k = 4$, $c = 0.1$ and $c_p = 0.08$.

service, given the cost c_m set (depicted at the top of the figure). The provider's benefit is maximized for $c_m = 0.2$ (yielding $B_p = 0.24$). Figure 8(b) depicts the agents' overall expected benefit as a function of c_m. Finally, Figure 8(c) depicts the agents' overall expected benefit if the communication provider sets a different c_m value (horizontal axis) and the agents compensate her for the loss due to switching to the new fee (i.e., the compensation is the difference between the provider's benefit with $c_m = 0.2$ and the new c_m according to 8(a)). As can be observed from the figure, given the option of side-payments the agents will offer

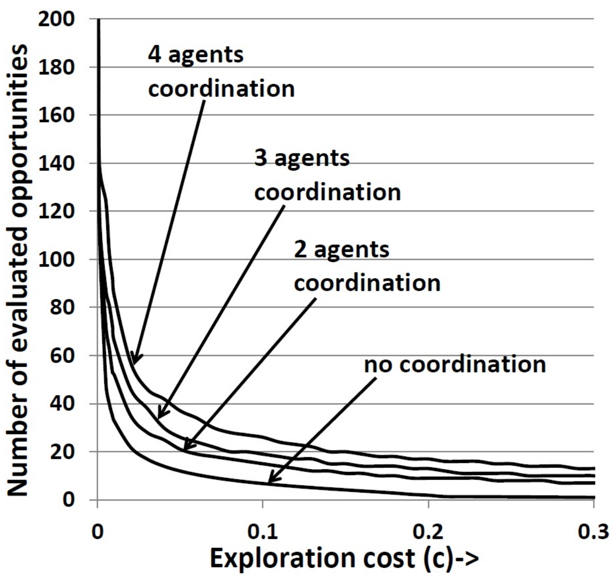

Figure 9. The maximum number of opportunities that will need to be evaluated in 99.9% of the cases (the 99.9*th* percentile) as a function of exploration cost, for four exploring agents.

the provider to use $c_m = c_p = 0.08$ and compensate her accordingly (see Section 3.5). This way, the agents' expected benefit increases from 1.173 to 1.2.

Before concluding, we note that the analysis was performed under the assumption that the agents are not limited by the number of opportunities they can potentially evaluate. While this is a standard assumption in costly exploration theories [27,30,26], it is very likely that even a moderate number of channels will do. Figure 9 depicts the 99.9*th* percentile of the number of opportunities that will be requested by the agents (i.e., the maximum number of opportunities that will need to be evaluated by any of the agents in 99.9% of the cases) as a function of the exploration cost, for the different options of dividing four exploring agents into a coordinated and non-coordinated subgroups.[10] The "no coordination" curve represents a scenario where all agents explore in isolation, with no coordination throughout the exploration. The "4 agents with coordination" relates to the case of full coordination. From the figure we observe that for most reasonable values of c, the number of opportunities that will need to be evaluated is relatively moderate (e.g., in this example when having 200 opportunities, in 99.9% of the cases none of the agents will require exploring an additional opportunity, according to the optimal strategy, for any cost of exploration $c > 0.0001$.[11] The importance of this observation is twofold: first, it justifies the use of a stationary set of thresholds as a solution to the problem, since the probability that the number of opportunities that the agents will want to evaluate will exceed the number

[10] It is notable that the coordination cost c_m has no effect over this figure since the curves depict all the different subgroups divisions and not just the best among them.

[11] Meaning that only for costs of less than 0.0001 more channels will be required. A cost of this magnitude is equivalent to 0.02%(!) of the mean of an opportunity's value (which is 0.5 in the uniform distribution case).

of available opportunities is, in most settings, negligible. Second, it reassures that the methods are applicable latency-wise, as the overall latency of the process is tightly correlated with the number of opportunities evaluated.

5. Experiments with Spectrum Sensing

In this section we demonstrate how the methods introduced and analyzed in the former sections can be applied in a real-world application. For this purpose we use the spectrum sensing application which also enables demonstrating how the analysis can be trivially adapted for the dual problem, i.e., when values represent "costs" or "expense", the value that can be used by all agents is the maximum among those found, and the goal is to minimize the overall expected sum.

5.1. Application

The application we use is spectrum sensing, i.e., the evaluation of different channels or frequencies one can potentially use as the physical layer for communication. Consider the case of two students who are working on a course assignment in the library[12] and need to transfer a file between their laptops using the wireless network. Both the establishment of communication with the University wireless network and the communication associated with the file transfer are battery consuming and the students are interested in draining their laptops' batteries as little as possible. Since the students are working on the same project, they are cooperative and their goal is to minimize the overall expected batteries power consumption which is equivalent to minimizing the overall expected time any of the laptops are engaged either in establishing communication with the network or actual file transmission. There are several available access points to the wireless network (e.g., different routers service set identifiers (SSIDs) or different channels that can be used with each router). These are the opportunities in our model, each associated with a different bandwidth (depending on the router's location, number of users currently using it, the current traffic level through the specific channel, noise, etc.). In many universities (e.g., at Bar-Ilan University which is used for our experiments) there is a single virtual access point (SSID) to the network, which can lead to any of the different physical access points (the choice of which is beyond the user's preferences or wills). In such cases, sequential attempts to reconnect to the network result in different physical connections, differing in the supported transmission bandwidth at the time of connection. The students thus will establish communication to the network according to some strategy and the effective transmission rate for transferring the file will be the minimum among the transmission rates of the two connections formed.

[12] While the example uses two students/agents, it can be generalized to any number of agents, k, as used in the analysis and results subsections that follow.

5.2. Optimal Strategies

Since the goal in the application we use in this section is to minimize the overall expected costs (i.e., the batteries draining) and the effective value used is the minimum bandwidth found, which is equivalent to the maximum time needed for transferring the file, the students actually engage with the dual problem to the one analyzed in this paper. The transition from the original problem to the dual one is straightforward as we show in the following paragraphs for the cooperative non-coordinated exploration analyzed in 3.2.

The optimal reservation value when disregarding others' findings (i.e., the method summarized in 3.1) in its overall expense minimizing version is given by [27]:

$$c = \int_{y=-\infty}^{r} (r-y)f(y)dy \tag{21}$$

Here, the exploration is resumed as long as the value obtained is *greater* than the reservation value.

When taking into consideration the value restriction, the optimal reservation value for the cooperative non-coordinated exploration that was analyzed in 3.2 can be derived from:

$$c=k\int_{y=-\infty}^{r} f(y)\Big(\int_{x=-\infty}^{\infty}(\max(r,x)-\max(y,x))\bar{f}(x)dx\Big)dy \tag{22}$$

where $\bar{f}(x)$ is given by:

$$\bar{f}(x) = \frac{d((F^{returned}(x))^{k-1})}{dx}, \tag{23}$$

where:

$$F^{returned}(x) = \begin{cases} 1 & x \geq r \\ \frac{F(x)}{F(r)} & x < r \end{cases} \tag{24}$$

Equations 22 and 23 can be obtained by following the proof given for Theorem 1 with the following changes. First, the expected value, EV, is calculated using the following modifications of Equation (3):

$$EV = \int_{y=r}^{\infty} EV f(y)dy + \int_{y=-\infty}^{r} f(y)\int_{x=-\infty}^{\infty} max(y,x)\bar{f}(x)dxdy \tag{25}$$

Second, the expected cost accumulated along the exploration is given by $\frac{c}{F(r)}$, as the resulting Bernoulli sampling process has a success probability of $F(r)$. The overall expected benefit of the system, EB, is a simple modification of (10) in a way that the expected cost is added to (rather than subtracted from) the expected value. Therefore, overall expected benefit is given by:

$$EB = k \cdot \left(\int_{x=-\infty}^{\infty} (x \cdot \frac{d(F^{returned}(x)^k)}{dx}) dx + \frac{c}{F(r)} \right) \quad (26)$$

5.3. Experimental Design

We used the Bar-Ilan University wireless network, to empirically extract the distribution of connections' transfer rates. For this purpose we used the IPERF software (`http://code.google.com/p/iperf/`) which is a network performance measurement utility that enables gathering information about the bandwidth of connections established. Each connection establishment with that tool takes 1 second. We ran a script which established connection to the network 20400 times, recording the bandwidth reported by the utility for each connection. Based on the data collected, we formulated the empirical probability distribution function of a random connection's bandwidth. In order to assure that re-querying the network gives a value according to that distribution (rather than having some dependency between subsequent queries) we randomly picked two samples of 50 consequent bandwidth readings, each starting at a random point at the stream of data. Then we run the Wilcoxon rank sum test, a non parametric test aiming to test the null hypothesis that the data in the two data sets are from the same distribution ([40], Chapter 2) and stored the p-value obtained. The null hypothesis that the two samples were taken from the same distribution function can be rejected if the p-value is substantially low (e.g., 0.01 or 0.05). We repeated the process 100 times, each time choosing two other random sequences of 50 consequent readings. The results show that the average p-value obtained is 0.22 (with a standard deviation of 0.31). This indicate that essentially there is a good reason to believe that indeed the sampling captured a steady distribution of bandwidth values, with a low dependency between subsequent readings (if at all). We repeated the process for different sample sizes, and obtained similar qualitative results: (0.25 average p-value, with a standard deviation of 0.31), (0.21 average p-value, with a standard deviation of 0.29) and (0.15 average p-value, with a standard deviation of 0.28) for 25, 100 and 500 sample sizes, respectively.

We assumed that the file that needs to be transferred is of size $10MB$. Therefore the connection bandwidth distribution could be expressed in terms of the transmission time distribution. This latter distribution was used as an input for extracting the expected-time-minimizing reservation values for the case where agents disregard others' findings (as summarized in 3.1) and the cooperative non-coordinated method (see 3.2). As emphasized before, the latter method is the one offering the least performance improvement and therefore was the one used for comparison. This allowed us calculating the expected overall time spent on communication establishment and transmission with the two compared methods.

5.4. Results

Figure 10(a) depicts the empirical bandwidth cumulative distribution function obtained based on the 20400 random connections established. The maximum connection bandwidth obtained was $2.46Mbps$ and the minimum was 0. Figure 10(b) depicts the expected-time-minimizing reservation values as a function of the num-

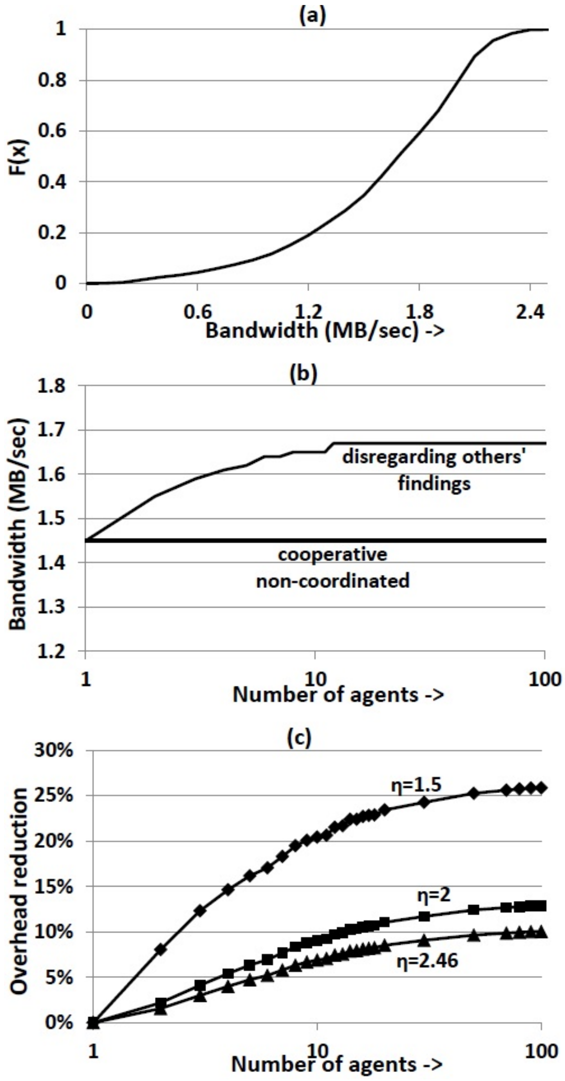

Figure 10. (a) The empirical bandwidth cumulative distribution function; (b) The expected overall time-minimizing reservation value as a function of the number of agents used (when having to transfer a $10MB$ file and $c = 1sec$); and (c) the percentage of exploration expected overhead improvement (reduction) for different theoretical bandwidth values when using the proposed cooperative non-coordinated method rather than disregard others' findings, as a function of the number of agents (when having to transfer a $10MB$ file and $c = 1sec$).

ber of agents, for each method (expressed in terms of the bandwidth, above which the exploration process should be terminated). As in Figure 5, and for the same explanation given there, in the cooperative non-coordinated exploration case the reservation value used by each agent increases as the number of cooperating agents increases, whereas in the other method it is not affected by the number of agents. The improvement achieved with the cooperative non-coordinated exploration is depicted in Figure 10(c), measured as the percentage of exploration expected overhead improvement (reduction) for different theoretical bandwidth values. Notice that the horizontal axis uses a logarithmic scale. The idea is that the improvement that can be achieved is always bounded by the time the file is transferred, and since the interval on which the bandwidth distribution is bounded then so does the expected change in the overall time. For example, since the maximum bandwidth that can theoretically be achieved is $2.46Mbps$, transferring the file itself will take at least 4.07 seconds, regardless of the exploration method used. Therefore, the goal is to aim how much we manage to improve beyond the file transfer time when using some theoretical bandwidth that one may obtain, denoted η. Given the theoretical bandwidth η, any time associated with the use of any of the methods beyond $10Mbps/\eta$ is actually an overhead associated with the exploration process. Given the expected times EB_1 and EB_2 when using the two method, respectively, the percentage of overhead reduction for a theoretical bandwidth η is given by: $1 - (EB_1 - 10/\eta)/(EB_2 - 10/\eta)$. We note that this calculation method is very strict, as it assumes the agents can transfer the file at rate η even without engaging with any exploration whatsoever. For Figure 10(c) we used three such theoretical bandwidth values: $\eta = 1.5Mbs$, $\eta = 2Mbps$ and the maximum theoretical possible value of $\eta = 2.4Mbps$. As can be observed from the figure, the percentage of overhead reduction achieved with the cooperative non-coordinated method is substantial. The actual improvement (overhead reduction) is actually substantially greater, since, as explained above, this comparison method assumes the agents can alternatively obtain an η bandwidth even without exploration, and also the non-coordinated method is the one offering the least performance improvement among those presented in this paper.

6. Related Work

In many multi-agent environments, autonomous agents may benefit from cooperating and coordinating their actions [34,43,44]. 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 [24,23,38,13]. Consequently, group-based cooperative behavior has been suggested in various domains [45,12,46]. 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 multi-agent systems (MAS) [25,38,18,16]. 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 [36,48,7], division of coalition payoffs [48] and enforcement methods for interaction protocols [29].

The problem of an agent engaged in exploration in a costly environment, seeking to maximize its long-term utility, is widely addressed in classical search theory (e.g., [47,41,27,35,26] 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 [37,17,21]. 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 [17,10,8]. Some even considered multi-agent cooperative exploration for multiple goals [37]. 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) [10,27]. Furthermore, none of these works considered costly coordination and its different aspects.

In prior work [33,32] we have considered multi-agent exploration in which one agent's exploration process is constrained by the findings of the other agents. Yet, in these models the exploration scheme was constrained, either in the sense that the agents are arbitrarily ordered and each agent can explore only after the other agents ordered before it have fully completed their exploration process (hence the coordination question becomes irrelevant) or by binding all agents to the same opportunity at any given exploration step (hence full coordination is mandatory). These constraints preclude the use of a hybrid exploration schemes of the type proposed in this paper, where some of the agents coordinate their exploration while others explore in isolation. These differences imply different exploration strategies and substantially complicates the analysis. Furthermore, since that prior work does not allow a partial coordination scheme, the modeling of an external self-interested communication provider and the resulting appropriate dynamics and side-payments aspects that are investigated in our paper becomes irrelevant there.

Multi-agent exploration can also be found in "two-sided" models (where dual exploration activities are modeled) [39,9,28]. 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.

Weakly related to our domain are the work on prize-collecting traveling salesman problems (PC-TSP) [4] and the graph searching problem (GSP) [22]. In PC-TSP one needs to pick a subset of the nodes to visit in order to minimize the total distance traveled while maximizing the total prize collected, given a graph with prizes associated with the different nodes. Despite the similar elements of the problem and its many variants [2,14], it substantially differ from our model in the sense that prizes are not uncertain and there is no cooperative exploration involved, hence findings are not affected/constrained by the performance of others. In the graph searching problem (GSP), the agent needs to find an item which location is associated with some probability function defined over the different nodes of the graph. The goal is indeed to minimize the expected cost, however the costs relate to the search itself and no tradeoff is given between findings and search costs. More broadly, our problem can be seen as part of the field of planning under uncertainty, hence related to Markov decision processes (MDP) [5,31] and decentralized Markov decision processes [6]. In these models the goal is to maximize the expected cumulative reward, which is also the objective in our case. Still, the use of MDPs for our cases is applicable only for the case where opportunity value is associated with a discrete probability function and the number of opportunities is finite. It does not provide a solution for a model of continuous probability distribution function and an infinite decision horizon. More important, our search-theory-based analysis and proofs result in a threshold-based solution which both simplifies the strategy and state representation and its complexity is substantially lesser than the one provided by MDPs.

7. Discussion and Conclusions

As common in MAS applications, coordination plays an important role in cooperative exploration, since it enables the agents to refine their exploration strategy based on the findings of others. Yet, since coordination is inherently costly, the agents should carefully reason about its extent of use. As the cost of coordination increases, the agents may find it more beneficial to have only some of them (if at all) coordinate their exploration. The analysis given in this paper enables the agents to compute their optimal (expected-benefit-maximizing) exploration strategy as a function of the coordination and exploration costs, for any extent of coordination used. This facilitates their decision regarding the amount of coordination they should apply.

The model and analysis extend traditional single-agent exploration models to a cooperative exploration one, according to which several agents 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. The analysis given in this paper proves that the optimal strategy that should be used by the agents is reservation-value based. Agents exploring in isolation (i.e., with no coordination with the others) will use a stationary reservation value, whereas those that use the coordination service will be using a state-based reservation value. The sequential nature of the exploration process used enables some level of separation in the analysis. The resulting set of equations that needs to be solved is actually based on each agent's best-response strategy given the distribution of the minimum value resulting from the exploration strategies of the other agents. In some cases this would require solving for an exponential divisions of the agents into sub-groups that need to be coordinated. In others, including the mostly common case where the cost of coordination is linear in the number of agents that consume it, we prove that the optimal exploration scheme is to have a sub-group of agents using the coordination service and all other agents exploring with no coordination with the others. In the latter case, thus, one only needs to extract the optimal exploration strategy for a number of divisions as the number of agents.

One important implication of the analysis, as illustrated in the former section, is that when including the strategic behavior of a self-interested coordination provider, the agents can benefit from the side-payments they make to the provider in exchange for charging a lower fee for the service. Doing so, the agents manage to increase the overall “net” benefit produced (having more agents coordinate their exploration) and take over a greater part of it. We note that if the provider could have priced the service in terms of bundles, it could have taken over all the benefit that could be generated through the use of coordination. Similarly, a mechanism for discounts could have been devised that would produce the same result. Still, such solutions are often inapplicable whenever the service is offered to many groups of agents of different sizes, and prices must similarly apply to all agents.

Finally, we note that while the model analyzed in this paper assumes that the coordination costs depend solely on the number of agents coordinated, there is much room for further research considering other types of coordination cost functions. For example, the self-interested communication provider may charge per communication message. In this case the exploration itself will be of a different structure, in comparison to the reservation-value based scheme used in this paper.

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. Arora and G. Karakostas. A $2+\epsilon$ approximation algorithm for the k-mst problem. In *Proceedings of the Eleventh Annual ACM-SIAM Symposium on Discrete Algorithms (SODA-2000)*, pages 754–759, 2000.

[3] J. Atlas and K. Decker. Coordination for uncertain outcomes using distributed neighbor exchange. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, pages 1047–1054, 2010.

[4] E. Balas. The prize collecting traveling salesman problem. *Networks*, 19(6):621–636, 1989.

[5] R. Bellman. A markovian decision process. *Indiana University Mathematics Journal*, 6(4):679–684, 1957.

[6] D. Bernstein, D. Givan, N. Immerman, and S. Zilberstein. The complexity of decentralized control of markov decision processes. *Mathematics of Operations Research*, 27(4):819–840, 2002.

[7] S. Breban and J. Vassileva. Long-term coalitions for the electronic marketplace. In *Proceedings of the E-Commerce Applications Workshop*, 2001.

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

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

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

[11] M. Chhabra, S. Das, and D. Sarne. Expert-mediated search. In *Proceedings of the 10th International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2011)*, pages 415–422, 2011.

[12] 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.

[13] P. Dutta and S. Sen. Forming stable partnerships. *Cognitive Systems Research*, 4(3):211–221, 2003.

[14] N. Garg. Saving an epsilon: a 2-approximation for the k-mst problem in graphs. In *Proceedings of the Thirty-Seventh Annual ACM Symposium on Theory of Computing (STOC-2005)*, pages 396–402, 2005.

[15] A. Grosfeld-Nir, D. Sarne, and I. Spiegler. Modeling the search for the least costly opportunity. *EJOR*, 197:667–674, 2009.

[16] B. J. Grosz and S. Kraus. Collaborative plans for complex group action. *Artificial Intelligence*, 86(2):269–357, 1996.

[17] 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.

[18] J. Kahan and A. Rapoport. *Theories of coalition formation.* L. Erlbaum Associates, 1984.

[19] B.-K. Kang. Optimal stopping problem with recall cost. *European Journal of Operational Research*, 117(2):222–238, 1999.

[20] B.-K. Kang. Optimal stopping problem with double reservation value property. *European Journal of Operational Research*, 165(3):765–785, 2005.

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

[22] E. Koutsoupias, C. H. Papadimitriou, and M. Yannakakis. Searching a fixed graph. In *Proceedings of the 23rd International Colloquium on Automata, Languages and Programming (ICALP-1996)*, pages 280–289, 1996.

[23] S. Kraus, O. Shehory, and G. Taase. Coalition formation with uncertain heterogeneous information. In *Proceedings of the Second International Conference on Autonomous Agents and Multi-agent Systems (AAMAS 2003)*, pages 1–8, 2003.

[24] 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.

[25] 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.

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

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

[28] 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.

[29] P. Michiardi and R. Molva. Analysis of coalition formation and cooperation strategies in mobile ad hoc networks. *Ad Hoc Networks*, 3(2):193–219, 2005.

[30] P. Morgan. Search and optimal sample sizes. *Review of Economic Studies*, 50(4):659–675, 1983.

[31] M. L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* Wiley-Interscience, 1994.

[32] I. Rochlin, D. Sarne, and M. Laifenfeld. Coordinated exploration with a shared goal in costly environments. In *Proceedings of the ECAI 2012 - 20th European Conference on Artificial Intelligence. Including Prestigious Applications of Artificial Intelligence (PAIS-2012) System Demonstrations Track*, pages 690–695, 2012.

[33] 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.

[34] A. Rosenfeld, G. Kaminka, S. Kraus, and O. Shehory. A study of mechanisms for improving robotic group performance. *Artificial Intelligence*, 172(6-7):633–655, 2008.

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

[36] 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.

[37] 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.

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

[39] R. Shimer and L. Smith. Assortative matching and search. *Econometrica*, 6(2):343–370, 2000.

[40] S. Siegel. *Non-Parametric Statistics for the Behavioral Sciences.* McGraw-Hill, 1956.

[41] L. Smith. Frictional matching models. *Annual Reviews in Economics*, 3(1):319–338, 2011.

[42] 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.

[43] P. Stone, G. Kaminka, S. Kraus, and J. Rosenschein. Ad hoc autonomous agent teams: Collaboration without pre-coordination. In *Proceedings of the Twenty-Fourth Conference on Artificial Intelligence*, pages 1504–1509, 2010.

[44] P. Stone and S. Kraus. To teach or not to teach? decision making under uncertainty in ad hoc teams. In *Proceedings of the Ninth International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, pages 117–124, 2010.

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

[46] 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.

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

[48] 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.

8. Nomenclature

Notation	Meaning
k	The number of individual agents that engage in the cooperative exploration (the group's size)
$f(x)$, $F(x)$	The probability density function and cumulative distribution function from which the values of the opportunities agents explore are drawn
c	The exploration cost of an opportunity
v^*	The minimum among the best results obtained in any of the individual exploration processes
$c_m(j)$	The cost of supplying j agents with coordination
B_p	The communication provider's benefit
$c_p(j)$	The provider's cost of providing coordination between j exploring agents
$\bar{f}(x)$	The probability distribution function of the minimum among the best values obtained by $k-1$ other agents in the non-coordinated optimal exploration
r	The optimal reservation value in a non-coordinated exploration
EV	The expected value with which the agents end up
EB	The system's overall expected benefit
$F^{return}(x)$	The cumulative distribution function of the minimum value obtained throughout an agent's exploration
V	The system's state represented as a vector (i.e., $V = (v_1, .., v_k)$) where v_i $(1 \leq i \leq k)$ is the best value found so far by agent A_i
$S(V)$	The agents' strategy, mapping a state V to a decision to resume exploration of agent A_i or terminate exploration
V_{min}	The minimum value in V (i.e., $V_{min} = min(v_1, .., v_k)$)
$r(V)$	The reservation value to be used when in state V
$\sigma(V, y)$	The new state to which the system transitions, if it was initially in state V, after the agent associated with V_{min} has obtained a value y in its exploration
$EB(V)$	The system's expected benefit, onwards, when in state V
k_i	The number of individual agents of the i-th subgroup exploring with coordination
$r_i(V)$	The reservation value function used by the agents of the i-th subgroup, upon reaching a state $V = \{v_1, ..., v_{k_i}\}$ in its coordinated exploration
$f_i^c(V, x)$	The probability distribution function of the minimum among the best values obtained by the k_i agents of the i-th subgroup exploring with coordination, given that they start from state V
$F_i^c(x)$	The probability that the agents in the i-th subgroup end up with a minimum value (among the best individually found) of x or below
$\bar{f}_i^c(x)$	The probability distribution function of the minimum of the best findings of all agents other than those in the i-th subgroup
$\bar{f}^c(x)$	The probability distribution function of the minimum of the best findings of all agents, except for one single agent among those that explore in isolation
$EV_i(V)$	The expected minimum value with which the agents end up given that the agents of the i-th subgroup are in state V and no other a priori information regarding the findings of the other agents
$EC_i(V)$	The expected overall cost of the i-th subgroup if using a reservation value function $r_i(V)$
EC	The expected overall cost of all k agents