Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence

Fair Information Sharing for Treasure Hunting

Yiling Chen
Harvard SEAS
yiling@seas.harvard.edu

Kobbi Nissim
Ben-Gurion University and Harvard CRCS
kobbi@cs.bgu.ac.il

Bo Waggoner
Harvard SEAS
bwaggoner@fas.harvard.edu

Abstract

In a search task, a group of agents compete to be the first to find the solution. Each agent has different private information to incorporate into its search. This problem is inspired by settings such as scientific research, Bitcoin hash inversion, or hunting for some buried treasure. A social planner such as a funding agency, mining pool, or pirate captain might like to convince the agents to collaborate, share their information, and greatly reduce the cost of searching. However, this cooperation is in tension with the individuals' competitive desire to each be the first to win the search. The planner's proposal should incentivize truthful information sharing, reduce the total cost of searching, and satisfy fairness properties that preserve the spirit of the competition.

We design contract-based mechanisms for information sharing without money. The planner solicits the agents' information and assigns search locations to the agents, who may then search only within their assignments. Truthful reporting of information to the mechanism maximizes an agent's chance to win the search. ϵ-voluntary participation is satisfied for large search spaces. In order to formalize the planner's goals of fairness and reduced search cost, we propose a simplified, simulated game as a benchmark and quantify fairness and search cost relative to this benchmark scenario. The game is also used to implement our mechanisms. Finally, we extend to the case where coalitions of agents may participate in the mechanism, forming larger coalitions recursively.

1 Introduction

A group of selfish pirates land on a forsaken island in search of a hidden treasure, an indivisible item of inestimable value. Each pirate has gathered limited information – a personal map marking certain locations on the island where the treasure might be located. Every day, each pirate can dig in a single location; whoever finds the treasure first will keep it forever. The pirate captain knows that, if only the pirates would share their information, many days of useless digging could be averted. If only she, as the wise and trusted leader, could convince the pirates to lend her their maps, then she could pool the collective knowledge and assign digging locations to minimize wasted effort. But can she assign locations in a way that is fair and just? And equally important, can she convince the pirates that it is in their best interests to all agree to give her their maps and abide by her assignments?

Our story abstracts settings where agents with heterogeneous information compete to solve a search problem. An example is when different research labs try to locate a gene corresponding to a genetic disease, and credit is only given to the first discoverer. Each of the researchers begins their search based on prior knowledge they acquired. Combining the researchers' prior information could speed up discoveries and reduce wasted effort.

This tension underlies the difficulties of *cooperation in a competitive environment*. A solution to the competing search problem must take into account many factors: incentives (agents must want to report accurate information); fairness (rewarding agents based on the progress made toward finding the answer to the search problem); and welfare (it should improve on the status quo by shortening the search).

We consider a basic setting where an agent's information consists of a set of possible locations where the solution ("treasure") may be found, and each location is equally likely to be the correct one. This simple model does not capture cases where agents have complex distributional beliefs. However, this setting already raises many interesting questions and difficulties. We believe that it can highlight the tension between cooperation and competition in situations such as the scientific credit example, although it may not capture cases with complex information structures.

A scenario where the assumptions of our model fit reality more closely is the Bitcoin digital currency protocol. The process of "mining" or creating new Bitcoins requires inverting a cryptographic hash function: We begin with a target output and some large set of possible inputs and search until we find the input that hashes to the output. Many miners may be searching in parallel to find the preimage first; they each have a list of possible values they have not yet tried; and each input is (approximately) equally likely. Although we will not suggest that our mechanisms should be directly applied to Bitcoin in practice, the example shows that the simple treasure-hunting model can already closely match some real-world settings. Indeed, "mining pools" or groups of cooperative Bitcoin searchers exist in practice in order to save unnecessary trials, save time, and improve their probability of winning. Our approaches are inspired by the same motivations.

1.1 Our approach: contract-signing mechanisms without monetary transfer

Our goal is to design mechanisms that help competing agents share their information. Our mechanisms are *contract-based* in the sense that agents first sign a "contract" saying the outcome of the mechanism – assignments to different agents of subsets of the (relevant part of the) search space – would be binding.[1] Only then, agents report their sets to the mechanism which computes and reveals the subset allocated to each agent.

Our mechanisms are implemented without monetary transfer. This increases their potential applicability to settings where the assumptions of monetary or transferable-utility mechanisms, such as quasilinearity of utility and no-budget assumptions, may not hold. For instance, in scientific research, it seems culturally implausible to suggest a money-based mechanism for aggregating knowledge.

One drawback of our mechanisms is that there exist situations where an agent might prefer not to participate. This could be undesirable for many settings such as in the scientific credit example. However, we show ϵ-voluntary participation where $\epsilon \to 0$ for large search spaces.

Other approaches. A body of literature with similar motivations to our work is that on cooperative game theory (CGT) (Osborne and Rubinstein 1994), which concerns coalition formation in games. The focus of CGT typically is on stability of a coalition and fairness in sharing value among members of a coalition. Our setting is superficially similar in that our mechanism forms "coalitions" of agents and we are interested in "fairness", but the treasure-hunting problem seems to clash with the usual cooperative game theory approach. Our setting is inherently *non-cooperative*. Partly, this is because bargaining needs to be done carefully, as private information, when revealed, has no more value; more importantly, this is because the pirates may misreport their information, and hence we must consider incentives and strategic behavior.

Trying to avoid making assumptions on how agents perceive other agents' information, we take a rather *agnostic* approach in modeling the information agents have. In our setting, agents are *not* required to form probabilistic beliefs about other agents' information. This is a weaker assumption than in classical Bayesian game settings, where it is assumed that the prior distribution of private information is common knowledge and agents must update according to this prior.

Design goals and benchmark. The first design goal is *incentives for truthful reporting*, so that the mechanism can correctly aggregate the agents' information. The second is *fairness*, which we interpret as preserving the spirit of the competition for searching for the treasure. An agent who has a good chance of finding the treasure without the existence of the mechanism should still have a good chance after the mechanism produces an assignment. The third is *welfare improvement*: the mechanism should reduce the total digging costs by combining agents' information.

In order to quantify the fairness and welfare goals, we introduce a hypothetical benchmark, the *simplified exploration game*. The idea is to imagine that all agents explore within their sets in a uniformly random order, regardless of the behavior of the others. In this simplified scenario, we can compute expected digging costs until the treasure is found, and also each agent's probability of finding the treasure first. Based on the benchmark we can set concrete quantifiable welfare and fairness goals.

A key insight of our approach to designing the mechanism is that we can use the simplified exploration game to get good incentives. Our mechanism takes the agents' reported sets and, based on these, computes the winning probability of each agent in the simplified exploration game. The set of possible treasure locations (obtained by intersecting the reported sets) is then divided by the agents in proportion to these computed winning probabilities. We show that this mechanism has good incentives regardless of what exploration strategy an agent might actually have planned to use.

Results summary. We first consider "one-shot" mechanisms: forming a coalition of the entire group of agents. We construct a one-shot contract-based mechanism and show that in this mechanism, to maximize winning probability, each agent should report her private information truthfully if all other agents report truthfully. Then, we prove the fairness and welfare properties of the mechanism. We also show that the mechanism satisfies ϵ-voluntary participation for $\epsilon \to 0$ as information sets grow large.

We then extend to a setting where several coalitions (each formed, say, by the one-shot mechanism) want to become one large coalition. We call these mechanisms "composable" because they can be used to recursively form larger and larger coalitions. We extend our approach to this setting and also begin an exploration of the dynamics that may result from the usage of such composable mechanisms.

Some proofs are omitted in order to save space and appear in the full version of the paper.

1.2 Related work.

It has been widely recognized that private information brings value and hence sharing of private information should be encouraged. (Kleinberg, Papadimitriou, and Raghavan 2001) draws on concepts in cooperative game theory to assign value to releasing private information in a few specific settings, including marketing surveys and collaborative filtering and recommendation systems. Interestingly, some recent work takes an opposite view, arguing that sometimes sharing less information improves social welfare or other objectives of the designer (Rochlin and Sarne 2014; Kremer, Mansour, and Perry 2013).

Our setting can model competition in scientific discovery. (Kleinberg and Oren 2011), (Kitcher 1990), and (Strevens

[1] We do not consider the question of enforcement in this paper. In the pirate story, the captain may behead the deviating pirates, a solution that we don't generally recommend.

2003) all model and study scientific development in the society. However, the strategic aspects of researchers in their models lie in the selection of research projects to work on; researchers who selected the same project compete independently. In particular, (Kleinberg and Oren 2011) study how to assign credits to projects so that the project selection behavior of self-interested researchers may lead to optimal scientific advances. Our setting essentially models a scenario with one project and instead of letting researchers independently compete on this project, we design mechanisms to allow them cooperate and share information while still competing with each others.

We use contract-based mechanisms to promote cooperation. Such approaches are common in other settings where some level of enforcement is necessary for incentive alignment. For example, (Wang, Guo, and Efstathiou 2004) design Nash equilibrium contracts to guarantee optimal cooperation in a supply chain game.

2 The Treasure Hunting Game

Let S (the island) be a finite set of locations, one of which is s^* (the treasure). There is a set N of agents who will be seeking the treasure, and $|N| = n$. Each agent i has as private information a *set* $S_i \subseteq S$, where it is guaranteed that $s^* \in S_i$. This immediately means that $s^* \in \cap_{i \in N} S_i$. We use S_N to denote the intersection $\cap_{i \in N} S_i$. The fact $s^* \in S_i$ for all $i \in N$ is common knowledge to all agents. We assume that each agent i believes that every element in S_i is equally likely to be s^*. We make no other assumptions on i's beliefs.[2]

Initially, the mechanism takes place: Each agent i reports a set $\hat{S}_i \subseteq S$ to the mechanism and receives a set $\Pi_i \subseteq S$ from the mechanism. i may *only* dig at locations in Π_i.

Subsequent is the *digging phase*, consisting of up to $|S|$ digging periods. In each period, each agent i can "dig" at one location $s \in \Pi_i$ of his choice. It is assumed that an agent will not dig in the same location twice. The digging phase ends immediately after the first period in which an agent digs at s^*. We assume that each agent wishes to maximize her probability of being the one to win the treasure.

It is assumed above that agents only dig at locations in their assigned set Π_i. This follows if agents agree beforehand to abide by the outcome of the mechanism and there is some manner of enforcing that they do so. Thus, we call the above procedure a *contract-signing mechanism*. We do not consider how the contract is enforced in this paper, but assume there exists a manner of enforcement.

Desiderata of the Mechanism. In the treasure-hunting scenario, the pirate captain wishes to satisfy three objectives:

- **Incentives**. The pirates should prefer to report all their information to the mechanism truthfully so that it can correctly aggregate.
- **Fairness**. The mechanism should be impartial among the agents and reward each according to the information he provides.
- **Welfare**. The mechanism should reduce the amount of wasted searching.

We formalize the desired incentive property by requiring that each agent maximize their probability of finding the treasure by reporting their information truthfully (assuming that others are not misreporting). This probability is over any randomness in the mechanism and over the randomness of the treasure location (recall that each pirate initially believes that it is uniformly distributed in S_i).

The fairness and welfare goals are more subjective. To meet them, the captain must answer the questions: What do we mean by "fair"? And how can we quantify "welfare" or reduced digging cost when we do not know what would have happened without our mechanism? (Perhaps some lucky pirate would have found the treasure on the first day!)

To answer both of these questions, we next define a simplified exploration game. This game will serve as a "benchmark" for fairness and welfare; the captain can compare her mechanism to what would happen in the benchmark game. We will also use this game as the basis for our proposed mechanism.

2.1 Formalizing Fairness: The Simplified Exploration Game

The *simplified exploration game* is defined as follows. We emphasize that the game is hypothetical and is not actually played by the agents. To emphasize this difference, we describe the game as being "simulated", say on a computer or as a video game with artificial players. In the game, each simulated player has a subset S_i of the island. The player chooses a permutation of her set S_i *uniformly at random*. This is the order in which the simulated player will dig in her set. Then, a simulated treasure location is drawn uniformly at random from the intersection S_N of the sets. Then, there is a sequence of simulated digging periods; in each period, each player "digs" at the next location in her chosen permutation. (In the simulation, this corresponds to simply checking whether the next location in the permutation is equal to the randomly drawn treasure location.) The simulation ends in the first period where some player simulates a dig at the simulated treasure location; this player wins the game. (Ties are broken uniformly at random.)

We next describe how the simplified exploration game can be used by the pirate captain as a benchmark for her subjective goals. In Section 3, we show how the captain can actually use the game to construct a mechanism.

- **Benchmark for fairness**: A mechanism can be considered fair if a pirate's chance to win the treasure under the mechanism matches his chance to win in the simplified exploration game. Intuitively, the simplified game is fair because *(a)* it rewards players for the value of their information: Players with smaller sets (better knowledge of the treasure) are more likely to win; *(b)* it rewards players *only* for the value of their information: A player cannot "jump ahead" of a better-informed opponent by em-

[2] For a concrete example model that implies such beliefs, suppose that the treasure is uniformly distributed on the island. Each agent receives as a signal a set of locations containing the treasure location, and updates to a posterior belief that the treasure is uniform on this set.

ploying some complex strategy; and *(c)* it preserves the competitive aspect of the treasure-hunting game: A player with high chances of winning in the game is guaranteed a high chance of winning under the mechanism, so he does not feel that the mechanism unfairly diminished his chance of winning.

- **Benchmark for welfare**: The welfare improvement of a mechanism is the difference in total expected exploration cost (number of locations searched) under the mechanism and in the simplified exploration game. (The expected digging cost for the mechanism is computed by assuming the treasure is uniformly random in the intersection and that each pirate explores her assignment Π_i in an arbitrary order.) This gives the captain a concrete measure of the mechanism's improvement. She can interpret this measure as saying something about the improvement the mechanism makes in real life, depending on how closely she thinks the simplified exploration game matches what would have happened without the mechanism.

3 One-Shot Mechanisms

In this section, we consider a *one-shot* setting, where all agents arrive and simultaneously participate in the mechanism. In Section 4 we will extend the discussion to the case where subsets of the agents have formed coalitions, and may wish to form even larger coalitions.

We propose that the captain utilize the simplified exploration game as a basis for a mechanism. The idea is to ask each pirate to report a set S_i, then consider the simplified exploration game where each pirate corresponds to a player. Then allocate digging locations according to performance in this simulated game.

More specifically, our primary mechanism for the one-shot setting is Mechanism 1, which proceeds as follows. First, all agents sign contracts agreeing to search only within their assigned location. Then, each agent i reports a subset $\hat{S}_i$ of the island to the mechanism. The mechanism computes the intersection $\hat{S}_N$ of the reports and assigns each element of the intersection independently at random according to the winning probabilities of the agents (with sets $\hat{S}_i$ in the simplified exploration game. Then, agents may dig only within their assigned subsets. (In particular, if the intersection is empty or the entire intersection is searched without discovering the treasure, agents are still not allowed to search elsewhere.)

We can imagine other allocation rules that use the winning probabilities from the simplified exploration game: for example, assigning locations deterministically with the number of locations proportional to the winning probabilities. So we can think of Mechanism 1 as giving a framework that can extend to any rule for dividing the intersection according to the winning probabilities. However, we do not explicitly consider these mechanisms and focus on Mechanism 1 for proving our results.

In the full version of the paper, we show how to efficiently compute these probabilities of winning for each agent and show the form of these probabilities, which is also useful for our results.

Mechanism 1: One-Shot Mechanism

Input: S_i for each agent i.
Output: A partition of $S_N = \cap_{i \in N} S_i$, with Π_i assigned to agent i.

```
set S_N = ∩_i S_i;
foreach agent i do
    compute i's winning probability p_i;
end
initialize each Π_i = ∅;
foreach location s ∈ S_N do
    let i be a random agent chosen with probability p_i;
    add s to Π_i;
end
output the sets Π_i for each i;
```

3.1 Results for one-shot mechanisms

Theorem 1. *In Mechanism 1, if other agents are reporting truthfully, then each agent i maximizes her probability of winning the treasure by reporting S_i truthfully.*

Proof. Under the mechanism, if other agents report truthfully, then agent i's probability of winning the treasure is exactly the probability (over the location of the treasure and the randomness of the mechanism) that the treasure location s^* is in i's assigned set Π_i. Thus, i prefers to report the set that maximizes this probability. We need to show that S_i is this set.

Some preliminaries: Denote agent i's report to the mechanism by $\hat{S}_i$, and fix the reports of agents except i to be truthful.[3] Denote the intersection of the reports by $\hat{S}_N$ and the probabilities of winning computed by the mechanism by $\hat{p}_i$ for each player i. Using this notation we get that $\Pr[i \text{ wins (when } i \text{ reports } \hat{S}_i)]$ is equal to $\Pr[s^* \in \hat{S}_N] \cdot \hat{p}_i$.

Let $MIN = \min_i |\hat{S}_i|$ be the smallest reported set size. In the full version of the paper, it is shown that the probability that i wins the treasure in the simplified exploration game given the reported sets can be written as

$$\hat{p}_i = \sum_{x=1}^{MIN} \frac{1}{|\hat{S}_i|} f_i(x) \tag{1}$$

where $f_i(x)$ is a probability that does *not* depend on $|\hat{S}_i|$, but only on the reports $\hat{S}_j$ for $j \neq i$. In particular, if $|\hat{S}_i|$ is not the unique smallest-sized set, then MIN does not depend on $\hat{S}_i$, and $\hat{p}_i$ is proportional to $|\hat{S}_i|$.

The proof proceeds as follows: We will show that for any fixed report $\hat{S}_i$, adding any location $s \notin S_i$ to $\hat{S}_i$ decreases this probability; and removing any location $s \in S_i$ from $\hat{S}_i$ decreases this probability. This will show that i's winning probability is maximized by reporting $\hat{S}_i = S_i$. Intuitively, the first case hurts i because she reports an unnecessarily

[3]To see why we cannot achieve a "dominant strategy" type of solution, suppose that all agents but i have committed to not reporting location $s \in S$, even if it is in their sets. Then s will not be in the intersection. So i is strictly better off by omitting s from her report, even if $s \in S_i$.

large set and thus unnecessarily decreases her probability of winning. In the second case, i obtains a higher probability of finding the treasure first in the simplified exploration game, but this is at least balanced out by the chance that the treasure was in the omitted location s (in which case it will not be in the intersection and nobody will get it).

Adding a location to $\hat{S}_i$. Let $s \notin S_i, \hat{S}_i$. Add s to $\hat{S}_i$ and use a prime symbol to denote the results of the change: $\hat{S}'_i = \hat{S}_i \cup \{s\}$; $\hat{S}'_N$ is the intersection when i reports $\hat{S}'_i$ rather than $\hat{S}_i$, fixing all other reports to being truthful; and $\hat{p}'_i$ is the computed probability for i to win in this case. Then, as the chance of s^* being in the intersection has not changed,

$$\begin{aligned}\Pr[i \text{ wins (when } i \text{ reports } \hat{S}'_i)] &= \Pr[s^* \in \hat{S}'_N] \cdot \hat{p}'_i \\ &= \Pr[s^* \in \hat{S}_N] \cdot \hat{p}'_i.\end{aligned}$$

If $|\hat{S}_i|$ is not the unique minimum-size set among all reports, then as discussed above, $\hat{p}_i$ is proportional to $\frac{1}{|\hat{S}_i|}$, so $\hat{p}'_i = \hat{p}_i \frac{|\hat{S}_i|}{|\hat{S}'_i|} = \hat{p}_i \frac{|\hat{S}_i|}{|\hat{S}_i|+1} < \hat{p}_i$. If it is the unique minimum-size set, *i.e.* $|\hat{S}_i| = MIN$ and $|\hat{S}_j| > MIN (\forall j \neq i)$, we still have $\hat{p}'_i < \hat{p}_i$. To see this, note that, in the formula for $\hat{p}_i$ where $|\hat{S}_i| = MIN$, the sum from $x = 1$ to MIN divided by $|\hat{S}_i| = MIN$ is an *average* over the values $f_i(x)$; and the same is true when $|\hat{S}'_i| = MIN$. However, this average can only decrease by including an additional term, because the terms are strictly decreasing (they are the probability of winning given on step x, as shown in the full version of the paper, and this must be strictly decreasing in x since any exploration order of the agents $j \neq i$ that allows i to win on day $x + 1$ also allows i to win on day x). So in either case, the probability that i wins when reporting $\hat{S}'_i$ is smaller than when reporting $\hat{S}_i$.

Removing a location from $\hat{S}_i$. Let $s \in S_i, \hat{S}_i$. Remove s from $\hat{S}_i$ and again use a prime symbol to denote the change. If $|\hat{S}_i| \neq MIN$, then analogously to above, $\hat{p}'_i = \hat{p}_i \frac{|\hat{S}_i|}{|\hat{S}'_i|} = \hat{p}_i \frac{|\hat{S}_i|}{|\hat{S}_i|-1}$. If $|\hat{S}_i| = MIN$, then we still have $\hat{p}'_i \leq \hat{p}_i \frac{|\hat{S}_i|}{|\hat{S}_i|-1}$, because we have the same multiplicative factor change and we are also summing over fewer terms. Meanwhile,

$$\begin{aligned}\Pr[s^* \in \hat{S}'_N] &= \Pr[s^* \in \hat{S}'_N \mid s^* \in \hat{S}_N] \cdot \Pr[s^* \in \hat{S}_N] \\ &= \frac{|\hat{S}_i| - 1}{|\hat{S}_i|} \cdot \Pr[s^* \in \hat{S}_N].\end{aligned}$$

Hence, the probability that i wins when reporting $\hat{S}'_N$ is

$$\begin{aligned}\Pr[s^* \in \hat{S}'_N] \cdot \hat{p}'_i &\leq \frac{|\hat{S}_i| - 1}{|\hat{S}_i|} \Pr[s^* \in \hat{S}_N] \cdot \hat{p}_i \frac{|\hat{S}_i|}{|\hat{S}_i| - 1} \\ &= \Pr[s^* \in \hat{S}_N] \cdot \hat{p}_i \\ &= \Pr[i \text{ wins (when } i \text{ reports } \hat{S}_N)].\end{aligned}$$

□

It is worth emphasizing that the incentive property is not compared to any sort of benchmark; it is an absolute property of the mechanism itself. For instance, even if an agent disliked the exploration game benchmark or disagreed that the mechanism satisfied good fairness properties, that agent would still agree that her probability of winning is maximized by reporting her set truthfully.

We next consider the desirable properties of fairness and welfare, as compared to the benchmark of the simplified exploration game.

Theorem 2. *Mechanism 1 satisfies fairness: the probability for an agent to win the treasure under the mechanism is equal to her probability of winning in the simplified exploration game.*

Proof. Immediate from the construction of the mechanism: The treasure is in some location s^* in the intersection S_N, and this location is assigned to player i with probability p_i, where p_i is her probability of winning the simplified exploration game. □

For welfare, the goal is to quantify the decreased exploration costs under the mechanism as compared to the benchmark. Specifically, we count the number of "digs" that take place, in expectation over the randomness of the mechanism and of the treasure location. For instance, if all n players dig on day 1, and then $n - 3$ players dig on day two, then $2n - 3$ "digs" have taken place. To measure the improvement, we focus on the parameter R which measures the potential "gain from cooperation". R is the ratio of the smallest agent set size to the size of the intersection. For instance, if every agent has a set of size 300, but by pooling their information they reduce their sets to just a size of 30, then $R = 10$.

Theorem 3. *Mechanism 1 satisfies the following welfare properties:*

1. $\mathbb{E}[\textit{\# digs with mech.}] \leq \mathbb{E}[\textit{\# digs in simp. exp. game}]$.
2. *Let* $R := \frac{\min_i |S_i|}{|S_N|}$*; then*

$$\frac{\mathbb{E}\left[\textit{\# digs with mech.}\right]}{\mathbb{E}\left[\textit{\# digs in simp. exp. game}\right]} \leq \frac{1}{2\frac{n}{n+1}R}\left(1 + \epsilon\right),$$

where $\epsilon = \epsilon(n, R, |S_N|) \to 0$ *as* $\frac{|S_N|}{n} \to \infty$.

To interpret the final result, note that, for large set sizes, then for two agents the ratio is approximately $\frac{1}{R}$. This means that (for instance) if both agents' sets are 10 times the size of the intersection, then the mechanism gives about a ten-fold improvement in digging cost. As the number of agents n also increases, the ratio approaches $\frac{1}{2R}$.

Voluntary Participation. One drawback to our mechanism is that it does not always satisfy voluntary participation, meaning that there are scenarios where an agent might rather not participate while all other agents do participate. This would not be a concern in many settings where participation is mandatory; for instance, all of the pirates vote on whether to implement a mechanism, and once the decision is made, all must participate together. But in many settings it is a desirable property. We show that the benefit from not participating is bounded by an ϵ increase in the probability of winning, where $\epsilon \to 0$ as the search space grows large. We assume for the theorem that an agent who does not participate explores uniformly at random. It may be of note that the MIN in the theorem statement is over all sets besides i's, so a single very well-informed agent is still incentivized to participate when others' sets are large.

Theorem 4. *There is an implementation of Mechanism 1 that satisfies ϵ-voluntary participation for $\epsilon \leq \frac{(n-1)(n-2)}{4}\frac{1}{MIN^2}$, where $MIN = \min_{j \neq i} |S_i|$. In particular, $\epsilon = 0$ for $n = 2$ and $\epsilon \leq \frac{n^2}{4MIN^2}$ for all n.*

4 Composable Mechanisms

In the previous section, we considered the case where all agents arrived and simultaneously joined a single "coalition". But what if some subsets of the agents have already met and formed coalitions? These coalitions might still be able to benefit from sharing information. This motivates our extension to "composable" mechanisms.

Our setting is exactly the same, except that entities wishing to participate in the mechanism may either be agents (as before) or *coalitions*. A coalition C is a set of agents along with an allocation rule for dividing the locations assigned to that coalition. Each agent i in the coalition has a set S_i and the intersection $\bigcap_{i \in C} S_i$ is denoted S_C.

Now, the mechanism should take in the coalitions $C_1, \ldots, C_m$ (we can think of individual agents as coalitions of size one) and output an allocation rule for dividing the intersection $S_N = \cap_{C_j} S_{C_j}$ among the agents. Then, before digging starts, this allocation rule is applied to produce a set of digging locations Π_i for each agent i; again, agents are contractually obligated to dig in their assigned sets. The goals are the same: good incentives (a coalition should maximize its probability of being allocated the treasure location by reporting S_C truthfully); fairness, and welfare. We next generalize the simplified exploration game and construct a mechanism that satisfies a corresponding notion of fairness.

4.1 Defining a Fair Mechanism

The simplified exploration game is generalized as follows (we can think of this as a "less-simplified exploration game"). First, we simulate each coalition C dividing its intersection S_C among its agents according to its allocation rule, which may be randomized. Lone agents can be interpreted as coalitions of size one who assign their entire set to themselves. Next, a simulated treasure location s^* is chosen uniformly at random from the grand intersection S_N of all sets. Finally, each agent picks a uniformly random permutation of her assigned set and explores in that order; the first to find the treasure wins (ties broken uniformly at random).

This exploration game extends the notion of fairness in the natural way. We will similarly use this exploration game as the basis for our composable mechanism, Mechanism 2. In analogy with Mechanism 1, we assign digging locations to coalitions randomly according to their probability of winning the "less-simplified" exploration game (more specifically, the probability that one of their members wins the game).

We do not know of a polynomial-time computable closed-form expression for the winning probabilities of the less-simplified exploration game. However, we still have two options for implementing Mechanism 2. First: For each location to be assigned, we simulate the exploration game once and assign that location to the winner. Second, we can estimate the winning probabilities of each agent by simulating the game many times, as mentioned in the single-shot case; a coalition's winning probability is the sum of its agents'.

Mechanism 2: Composable Mechanism

Input: A set of coalitions $C_1, \ldots, C_m$.
Output: A coalition N whose members are the union of the members in the input coalitions.

```
set S_N = ∩_j S_{C_j};
output N, whose set is S_N and whose allocation rule is as
follows:;
foreach coalition C_j do
    foreach agent i ∈ C_j do
        set or approximate p_i using the simulated exploration
        game;
    end
end
initialize each Π_i = ∅;
foreach s ∈ S_N do
    let i be a random agent chosen with probability p_i;
    add s to Π_i;
end
```

4.2 Incentives for the Composable Mechanism

The composable mechanism also satisfies our desired incentive property, that truthful reporting maximizes probability of winning. In addition, we also briefly consider incentives for coalition formation.

Theorem 5. *In Mechanism 2, given that other coalitions are reporting their sets truthfully, each coalition C maximizes the probability that an agent in that coalition finds the treasure by reporting its true set S_C.*

A note on coalition formation. Two pirates are discussing their treasure-hunting strategies on the ship as it sails to the island. They realize that they would be better off sharing information, so they decide to form a coalition using a fair contract-signing mechanism (say, Mechanism 1). Later that evening, while scrubbing the decks, they meet a group of three pirates who have already formed a coalition of their own. The two coalitions talk things over and agree to merge to form a five-person coalition, using Mechanism 2. And the process continues.

Since Mechanism 2 takes coalitions as input and produces coalitions, it can be used recursively (*i.e.*, the input coalitions had originally formed using Mechanism 2, possibly from other coalitions, etc). We can think of the entire process as being described by a *formation tree*, where the leaves are individual agents and each node is a coalition. A node's parent, if any, is the coalition that the node joins.

This is primarily a direction for future work, and we do not explore this question in any depth, but just consider one initial question. Suppose we fix a formation tree and pick a single agent. Would that agent's choice be to join the tree earlier or later than they currently join? We show that they prefer to join as early as possible, up to a vanishing ϵ. The same holds for coalitions of agents.

Theorem 6. *Under Mechanism 2, entities always ϵ-prefer to join a formation tree earlier than they currently do. That is, for any fixed formation tree, a coalition decreases its winning probability by no more than ϵ if removed from its current parent node and attached to any node along a path from that parent to a leaf. (It may increase its winning probability arbitrarily.) ϵ can be bounded by the probability of a tie in the simplified exploration game, $\frac{n}{\min_{i \in N} |S_i|}$.*

5 Discussion and Future Work

The treasure hunting problem is one way to abstract the problem of *cooperation in competitive environments*. We identified the key goals in this setting as good incentives for truthful reporting (allowing information aggregation), fairness (preserving the spirit of the competition), and welfare (reducing wasted search costs). We initially constructed single-shot mechanisms for all agents to participate in, then "composable" mechanisms in which coalitions can merge to form larger coalitions.

This direction suggests the problem of *dynamics* of coalition formation over time. If agents can strategically form coalitions, but have incomplete information about others' information, how will they behave? How can a mechanism designer incentivize the formation of a simple, single grand coalition rather than fragmented strategic formation? There seem to be many potential avenues to explore this question.

Non-uniform distributions and Bayesian models. It is natural to raise the question of a non-uniform distribution on treasure locations, and the related (but separate) question of a Bayesian model of agent beliefs.

For a non-uniform distribution, one approach could be to "re-cut" the island into pieces of equal probability to recover a uniform distribution; but the pieces might not take the same time to explore, raising new challenges. In this light, the uniform distribution assumption might be interpreted as saying that probability of finding the treasure in a location is proportional to the work it takes to explore that location. Non-uniform distributions also raise the question of what to do if the designer does not know the distribution or if agents have differing or irreconcilable beliefs.

A Bayesian model of the treasure hunting problem would have the potential to address many different questions than the ones considered in this paper. It would require stricter assumptions than this paper: In a Bayesian game, agents must form beliefs about the knowledge and actions of others. We allowed agents to be agnostic as to others' information and digging strategies, not requiring (for instance) common knowledge of the information structure. (A Bayesian model in which the treasure is uniformly distributed over the island would be compatible with our assumptions, but would make stronger assumptions that we do not need.) However, the obvious benefit of a Bayesian model would be to consider more sophisticated information models and perhaps focus on strategic aspects of play.

One could apply the "simplified-game" approach in this paper to construct a "direct-revelation" Bayesian incentive-compatible mechanism: Ask each agent to report, not just their set S_i, but additionally a *strategy* for exploring the island. Simulate the exploration game using these reported strategies (rather than uniform random exploration as in this paper), and allocate states from the intersection according to winning probabilities. Alternatively, the mechanism could collect only reports of the sets S_i, attempt to compute a Bayes-Nash equilibrium on behalf of the players (or a correlated equilibrium), and simulate equilibrium strategies. Two challenges for this sort of approach are, first, how to model information (in particular, what the mechanism needs to know to aggregate reports or compute an equilibrium); and second, how to define and achieve fairness in the Bayesian setting.

Acknowledgments

The treasure hunt problem was suggested to us by Claudio Orlandi. We thank Claudio for initial discussions of the problem. We also thank the anonymous reviewers for feedback, ideas, and suggestions.

This work was partially supported by the National Science Foundation, under grant CCF-0953516. Any opinions, findings, conclusions, or recommendations expressed here are those of the authors alone. Research done while K.N. was visiting Harvard University, supported by NSF grant CNS-1237235, ISF grant (276/12), and a gift from Google, Inc.

References

Kitcher, P. 1990. The division of cognitive labor. *The Journal of Philosophy* 87(1):pp. 5–22.

Kleinberg, J., and Oren, S. 2011. Mechanisms for (mis)allocating scientific credit. In *Proceedings of the Forty-third Annual ACM Symposium on Theory of Computing*, STOC '11, 529–538. New York, NY, USA: ACM.

Kleinberg, J.; Papadimitriou, C. H.; and Raghavan, P. 2001. On the value of private information. In *Proceedings of the 8th Conference on Theoretical Aspects of Rationality and Knowledge*, TARK '01, 249–257. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Kremer, I.; Mansour, Y.; and Perry, M. 2013. Implementing the "wisdom of the crowd". In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, EC '13, 605–606. New York, NY, USA: ACM.

Osborne, M. J., and Rubinstein, A. 1994. *A course in game theory*. MIT Press.

Rochlin, I., and Sarne, D. 2014. Constraining information sharing to improve cooperative information gathering. In *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems*.

Strevens, M. 2003. The role of the priority rule in science. *The Journal of Philosophy* 100(2):pp. 55–79.

Wang, H.; Guo, M.; and Efstathiou, J. 2004. A game-theoretical cooperative mechanism design for a two-echelon decentralized supply chain. *European Journal of Operational Research* 157(2):pp. 372–388.