Implemeting the "Wisdom of the Crowd" *

Ilan Kremer† Yishay Mansour‡ Motty Perry§

29-July-12

Abstract

We study a novel mechanism design model in which agents arrive sequentially and each in turn chooses one action from a set of actions with unknown rewards. The information that becomes available affects the incentives of an agent to explore and generate new information. We characterize the optimal disclosure policy of a planner whose goal is to maximizes social welfare. One interpretation for our result is the implementation of what is known as the 'Wisdom of the crowds'. This topic has become more relevant during the last decade with the rapid adaptation of the Internet.

*We wish to thank Michael Borns for his invaluable editorial work.

†Ilan Kremer: Stanford University and the Hebrew University of Jerusalem, ikremer@stanford.edu.

‡Yishay Mansour: Tel Aviv University, mansour@tau.ac.il. This research was supported in part by the Google Inter-university center for Electronic Markets and Auctions, by The Israeli Centers of Research Excellence (I-CORE) program, (Center No. 4/11), by a grant from the Israel Science Foundation, by a grant from United States-Israel Binational Science Foundation (BSF), and by a grant from the Israeli Ministry of Science (MoS).

§Motty Perry: University of Warwick and The Hebrew University of Jerusalem, motty@huji.ac.il.

1 Introduction

The Internet has proven to be a powerful channel for sharing information among agents. In doing so it has become a critical element in implementing what is known as the 'Wisdom of the crowds'. Hence it is not that surprising that one of the most important recent trends in the new Internet economy is the rise of online reputation systems that collect, maintain, and disseminate reputations. There are now reputation systems for such things as high schools, restaurants, doctors, travel destinations, and even religious gurus, to name just a few. A naive view is that perfect information sharing through the Internet allows for optimal learning and support the optimal outcome. We argue that this is not the case and present here a first step toward a characterization of the optimal strategy to share information when agents behave strategically.

To examine these issues we study a novel mechanism design problem in which agents arrive sequentially one after the other and each in turn chooses one action from a fixed set of actions with unknown rewards. The agent's goal is to maximize his expected rewards given the information he possesses at the time of arrival. A principal, whose interest is in maximizing the social welfare, is the only one to observe all past outcomes and can affect the agents' choices by revealing some or all of his information. His problem then is to choose an optimal disclosure/recommendation policy taking into account that the agents' short-term goals are not always in line with welfare maximization.

To see why full transparency may not be optimal consider the following simple example. Suppose that there are two alternatives, and agents share common priors regarding the two alternatives where μ_j represent the prior mean of alternative $j = 1, 2$. Each agent selects an action only once and suppose that once an alternative is visited its deterministic payoff x_j is realized, and there is no further uncertainty about its payoff. The question is how information is produced and shared among the different agents in order to

maximize the social welfare. Consider first the case in which agents cannot share their experience (in the context of the Internet this can be thought as the pre-Internet age). Assuming that $\mu_1 > \mu_2$, then all agents will choose the first alternative. Now suppose that there is perfect information sharing. The first agent still chooses the first alternative, and unlike before, he now reports his experience on the web-site so everyone can see. In case $x_1 < \mu_2$ the second agent will visit the second alternative and all other agents will choose the better alternative and the outcome is efficient. However, if $x_1 > \mu_2$ then all agents will choose the first alternative despite the fact that from a social perspective it is inefficient. There is a significant probability that the second alternative is much better.

The reason why perfect information sharing is not optimal is that it does not address the incentives of selfish agents, and thus does not allow for enough exploration. Agents, in our set up (as they are in the Internet economy), not only consume information but also produce information which in turn can be consumed by others. However, information is a public good and as such one needs to be careful in providing proper incentives for an agent to explore and produce new information.

The new 'Internet Economy' provides several related examples for which our model is relevant. Web site such as yelp.com, TripAdvisor.com and others try to collect information from users while making recommendations to these users. In a sense a manager of such a web site can be viewed as the social planner who implements what economists describe as the 'wisdom of the crowd'. As we argue in this paper the manager of these web sites is facing a non-trivial task as there is tension between gathering information from users while making recommendations to the same users.

An interesting example is a company called Waze-Mobile which has developed a GPS navigation software that is based on the wisdom of the crowd. Waze is a social mobile application providing free turn-by-turn navigation based on the live conditions of the road 100% powered by users. As many

drivers use this software, the more benefit it is to customers. When a customer log in to Waze with his smartphone, he continuously send information to Waze about his speed and location and this information, together with information sent by others, enable Waze to recommend to this driver as well as all other drivers an optimal route to their destination. But in order to provide good recommendation, Waze must have drivers in every possible route. Indeed as was described by Waze president and co-founder (see http://www.ustream.tv/recorded/21445754) they often recommend a driver a particular route even though (indeed exactly because) they do not have information about that route. The information transmitted by this driver is then used to serve better future drivers. But in order not to deter drivers from using the system, Waze must be very careful in how often they "sacrifice" drivers to improve the experience of others.

Finally consider the recent controversy over the health care report-card system. This system entails a public disclosure of patient health outcomes at the level of the individual physician. Supporters argue that the system gives providers powerful incentives to improve quality together with providing patients with important information. Skeptics counter that report cards may encourage providers to "game" the system by avoiding sick patients, seeking healthy patients, or both. We look at this problem from a different angle by asking how to optimally reveal the available information to maximize the social welfare taking into account the users incentives.

We next present in Section 2 the simplest possible model that enables us to study this problem. In the model the set of actions contains only two *deterministic* actions with unknown rewards. Then in Section 3 the principal's optimal policy is characterized. To this end we first provide a version of the *revelation principle* and then derive the optimal policy which is shown to be intuitive and simple. In the optimal policy agent one always choose the action with the higher mean and we denote his reward by r. If $r \in I_t$ then agent t is the first agent to whom the principal recommends to

try the other action while for all agent $t' > t$ the recommendation is the better of the two. The bulk of the analysis is devoted to fully characterize the sequence $\{I_t\}_{t\in T}$ which is shown to be a monotone partition of the real. As the number of agents increases the social welfare in the optimal policy converges to the first-best welfare in the unconstrained mechanism. Our focus is on the information channel and hence through out most of the paper we restrict our attention to the case in which the principal can only control the information that is available to agents and in particular is not allowed to use transfers. In section 4 we allow the principal to use monetary transfers in order to enhance incentives. It is shown that the essential properties of the optimal policy are unaffected by the introduction of transfers.

1.1 Related Literature

The literature on informational cascades which originated with the work of Bikhchandani, Hirshleifer, and Welch (1992) is probably the closest to the model presented here. An informational cascade occurs when it is optimal for an individual, who has observed the actions of those ahead of him, to follow the behavior of the preceding individual without regard to his own information. Our problem is different as we examine social planner who can affect the information received by each individual while implementing the optimal informational policy.

The agents in the model considered here are choosing from a set of two-armed bandits (see the classical work of Rothschild (1974)). But unlike the vast early work on the topic which was entirely about single-agent decision-making, our work is along the lines of the more recent works on strategic experimentation where several agents are involved, as in the work of Bolton and Harris (1999) and Keller, Rady, and Cripps (2005), to name just a few. The deviation from the single-agent problem is that an agent, in this multi-agent setting, can learn from experimentation by other agents. Information is therefore a public good, and a free-rider problem in experimentation naturally

arises. It is shown that because of free-riding, there is typically an inefficiently low level of experimentation in equilibrium in these models. In contrast, in our model, free-riding is not a problem as agents have only one chance to act, namely, when it is their turn to move. However, like in the above models, in our model information transmission plays an important rule as we let the planner choose what information to release and when. Again, our contribution is in approaching the problem from a normative, mechanism design point of view.

A related paper is Manso (2012) which studies an optimal contract design in a principal-agent setting in which the contract motivates the agent to choose optimally from a set of two-armed bandits. Yet, while in Gustavo's setup there is one agent who works for two periods, in our setup there are multi-agents who choose sequentially.

The planner in our model is not allowed to use monetary transfers as a tool to provide incentives. Mechanism design without monetary transfers has been with us from the early days when the focus of interest was the design of optimal voting procedures. One such model which also have the sequential feature of our model is Gershkov and Szentes (2009) in which a voting model is analyzed where there is no conflict of interest among voters and information acquisition is costly. In the optimal mechanism the social planner asks, at random, one voter at a time to invest in information and to report the resulting signal. In our model the order according to which agents arrive is given and known to every one. It is not difficult to see that if in our set up agents do not know their place in line then the first-best outcome is easily achieved. In recent years, the interest in this type of exercise has gone far beyond voting, as for example in the paper of Martimort and Aggey (2006) which considers the problem of communication between a principal and a privately informed agent when monetary incentives are not available. The paper by Kamenica and Gentzkow (2011) is very relevant to ours. They consider a sender-receiver game in which the sender is required to reveal

all the information she obtains but has control over the precision of this information. They show that choosing a fully informative signal might not be optimal and so is no information.

2 Model

We consider a binary set of actions $A = \{a_1, a_2\}$. The reward R_i of action a_i is deterministic, but ex-ante unknown. We assume that R_i is drawn independently from a continuous distribution D_i that is common knowledge and we let $\mu_i = E_{R_i \sim D_i}[R_i]$. Without loss of generality, we assume that $\mu_1 \geq \mu_2$.

There are T agents who arrive one by one, choose an action, and realize their payoff; they do not observe prior payoffs and only know their place in line. The planner observes the entire history and *commits* to a message (disclosure) policy, which in the general setup is a sequence of functions $M = \{\tilde{M}_t\}_{t=1,\ldots,T}$ where $\tilde{M}_t : H_{t-1} \to M_t$ is a mapping from the set of histories H_{t-1} with length $t-1$ to the set M_t of possible messages to agent t.

The goal of agent t is to maximize his expected payoff conditional on his information while the goal of the planner is to maximizes the expected average reward, i.e., $E[\frac{1}{T}\sum_{t=1}^{T} R_t]$. An alternative objective for the planner would be to maximize the discounted payoff, $E[\sum_{t=1}^{T} \gamma^t x_t]$, for some discounting factor $\gamma \in (0,1)$. We focus on the average payoff as it is more appropriate to our setup, but a similar result holds if the planner wishes to maximize the discounted payoff.

Before we proceed to characterize the optimal solution we note that one can generalize our model so that the distribution of payoffs does not have full support. The distribution does not even need to be continuous. These assumptions are made to simplify the exposition. However, it is important that while $\mu_1 \geq \mu_2$ there is a positive probability that the first action's payoff is lower than μ_2; that is, $\Pr(R_1 < \mu_2) > 0$; this holds when we assume full

support. If, instead, $\Pr(R_1 < \mu_2) = 0$, then all the agents will only choose the first action regardless of any recommendation policy. This follows as all agents are certain that the payoff of the first action exceeds the mean of the second action. In such a setup a planner will find it impossible to convince agents to explore.

3 Optimal Truthful Mechanism for two actions

Let us first give an overview of the mechanism and the proof. We start by providing a simple example that illustrates the main properties of the optimal mechanism. Then in Subsection 3.2 we present some few basic properties of incentive compatible mechanisms. In particular we establish a revelation principle version for our set-up where we show that without loss of generality, we can concentrate on recommendation mechanisms, that specify for each agent which action to perform (Lemma 1). We show that once both actions are sampled, the mechanism can always recommend the better action and stay incentive compatible (Lemma 2).

In Subsection 3.3 we explore the incentive compatible constraint of the agents, and show that any mechanism that is incentive compatible to the worse a prior action, is incentive compatible to both (Lemma 3). This simplifies the discussion to concentrate only on the incentives of the worse apriori action.

Subsection 3.4 develops the optimal mechanism. We first show that initially the optimal mechanism explores as much as possible (Lemma 4). We then show that any value of the better a priori action which are lower than the expectation of the other action, causes an exploration already by the second agent (Lemma 5). The main ingredient in our proof is that the lower realizations cause exploration before higher realizations (Lemma 6). Finally, there is some value of the better action, that realizations above it

cause the principal not to do any exploration.

This implies the optimal incentive compatible mechanism is rather simple. The principal explores as much as he can (given the incentive compatible mechanism) until a certain value (depending on T the number of agents) for which it does not perform any exploration.

3.1 Example

Consider a simple example in which the payoff of the first alternative, R_1, is distributed uniformly on $[-1, 5]$ while the payoff of the second alternative, R_2, is distributed uniformly on $[-5, 5]$. Assume also that T is large enough so it is optimal to test the two alternatives as early as possible and from then on to choose the better of the two.

Assume first what would happened in the case of full transparency. The first agent chooses the first action. The second agent would choose the second alternative only if the payoff of the first alternative is negative, $R_1 \leq 0$. Otherwise he and all the agents after him will choose the first alternative, an outcome which is suboptimal if T is large.

Now consider a planner who does not disclose R_1 but instead recommends the second alternative to the second agent whenever $R_1 \leq 1$. It is easy to verify that in this case, the second agent would follow the recommendation. The reason for this is that conditional on being recommended the second alternative he concludes that the expected value of the first alternative conditional on this recommendation is zero which is equal to the expected value of the second alternative. Based on our assumption of T being sufficiently large this implies that the outcome under this policy is more efficient than the one under full transparency as we will have more experimentation by the second agent. Hence, we can already conclude that full transparency is sub-optimal. But we can do even better.

Consider next the third agent where things become more interesting. Suppose that the planner policy is such that he recommends agent three to use

the second alternative if one of two cases occurs (I) the second agent has been recommended to test the second action $(R_1 \leq 1)$ and based on the expirience of the second agent the planner knows that $R_2 > R_1$, and, (II)- the third agent is the first to be recommended the second alternative because $1 < R_1 \leq 1 + x$. When calculating the benefit from choosing the second alternative agent three considers two cases:

I: $R_1 \leq 1, R_2 > R_1$: in this case the third agent is certain that the second alternative has already been tested by the second agent and was found to be optimal; this implies that $R_2 > -1$. When computing the expected gain conditional on this event, one can divide it into two sub-cases: I_a : $R_2 > 1, I_b : R_2 \in [k-1, 1]$. The probability of these two events (conditional on case I) are:

$$\begin{aligned} \Pr(I_a|I) &= \frac{\Pr(R_2 > 1, R_1 \leq 1, R_2 > R_1)}{\Pr(R_2 > 1, R_1 \leq 1, R_2 > R_1) + \Pr(R_2 \in [k-1, 1], R_1 \leq 1, R_2 > R_1)} \\ &= \frac{0.4 * 1/3}{0.4 * 1/3 + 0.2 * 1/3 * 1/2} = 0.8 \end{aligned}$$

$$\Pr(I_b|I) = 1 - \Pr(I_a|I) = 0.2$$

The gain conditional on (I_a) is: $E(R_2 - R_1|I_a) = E(R_2|R_2 > 1) - E(R_1|R_1 < 1) = 3 - 0 = 3$. The gain conditional on I_b is $E(R_2 - R_1|I_b) = E(R_2 - R_1|R_1, R_2 \in [-1, 1], R_2 > R_1) = 2/3$. Hence, the gain conditional on I is given by:

$$E(R_2 - R_1|I) = \frac{0.8 * 3 + 0.2 * 2/3}{0.8 + 0.2} = \frac{38}{15}$$

The relative gain from following the recommendation when we multiply by the probability of I is:

$$\Pr(I) * E(R_2 - R_1|I) = \frac{2}{2+x} * \frac{38}{15}$$

$II : 1 < R_1 \leq 1 + x$: Conditional on this case our agent is the first to test

the second alternative. The expected loss conditional on this event is

$$E(R_1 - R_2|II) = E\left[R_1|R_1 \in [1, 1+x]\right)] - E(R_2) = \frac{1+(1+x)}{2} - 0 = \frac{2+x}{2} .$$

When we multiply this by the probability of this event we get:

$$\Pr(II) * E(R_2 - R_1|II) = \frac{x}{2+x} * \frac{2+x}{2} = \frac{x}{2}$$

Equating the gain and the loss yields $x = 2.23$. This implies that if agent $t = 3$ is recommended the second action when $I : R_1 \leq 1$ and the planner has learnt that the second action is optimal or when $II : 1 < R_1 \leq 3.23$ he will be willing to follow the recommendation. The computation for the fourth agent is similar, and here we get that this agent will explore (i.e., be the first to test R_2) for the remaining values of R_1, i.e., $R_1 \in [3.23, 5]$. All the remaining agents are recommended the better of the two actions.

The rest of the paper is devoted to show how this logic can be extended to form the optimal policy and to show that the number of exploring agents is a constant, independent of the number of agents.

3.2 Preliminary

We start the analysis with two simple lemmas that, taken together, establish that it is possible without loss of generality to restrict attention to a special class of mechanisms in which the principal recommends an action to the agents, and once both actions are sampled, the better of the two is recommended thereafter. The first lemma is an application of the well-known *Revelation Principle* to our setup.

Definition 1 *A recommendation policy is a mechanism in which at time t, the planner recommends an action $x_t = a_j$ and it is incentive compatible for the agent to follow the recommendation, that is, $E[R_j - R_i|x_t = a_j] \geq 0$ for each $a_i \in A$.*

Note that our definition of a recommendation policy includes the requirement that it is incentive compatible.

Lemma 1 *For any mechanism M, there exists a recommendation mechanism that yields the same expected average reward.*

Proof: For an arbitrary mechanism M, let M_t^j denote the set of all messages that lead agent t to choose the action a_j and let $H_{t-1}^j = (M_t^j)^{-1}$ denote the corresponding set of histories that lead to a message from M_t^j. It follows that for each $m \in M_t^j$ we have $E[R_j - R_i|m] \geq 0$. Now, consider a recommendation mechanism that recommends action a_j whenever the history is in H_{t-1}^j. Note that this mechanism is also incentive compatible since for each $m \in M_t^j$ we have $E[R_j - R_i|m] \geq 0$. Since it results in identical choices by the agents it results in an identical payoff. □

The next lemma allows us to narrow further the set of mechanisms that we refer as the set of partition policies.

Definition 2 *A partition policy is a recommendation policy that is described by a collection of disjoint sets $\{I_j\}_{j=2}^{T+1}$. If $R_1 \in I_t$ then agent t is the first agent for whom $x_t = a_2$ and for all $t' > t$ we have $x_{t'} = \max\{a_1, a_2\}$. If $R_1 \in I_{T+1}$ then no agent will be recommended to use the second action.*

Lemma 2 *If Π is an optimal recommendation mechanism, then Π is a partition mechanism.*

Proof: Note first that since $\mu_1 \geq \mu_2$ the first agent would always choose the first action. Also, since the principal wishes to maximize the average reward, i.e., $E[\frac{1}{T}\sum_{t=1}^{T} R_t]$, it would always be optimal for him to recommend the better action once he has sampled both actions. Hence, for each agent $j \geq 2$ we need to describe the realizations of R_i that would lead the planner to choose agent j to be the first agent to try the second action. Clearly, recommending the better of the two actions will only strengthen the IC of the agent to follow the recommendation. □

A partition policy has two restrictions. The first is that it recommends to the first agent action a_1. This is an essential condition to be *IC*. The second is that once it has sampled both actions it recommends the better one. Clearly this is an essential property of being optimal. In what follows we will restrict our attention to partition policies.

3.3 Incentive-Compatibility (IC) Constraints

Agent t finds the recommendation $x_t = a_2$ incentive compatible if and only if

$$E(R_2 - R_1|\text{principal recommends } a_2) \geq 0 \ .$$

Note that this holds if and only if

$$\Pr(\text{principal recommends } a_2) * E(R_2 - R_1|\text{principal recommends } a_2) \geq 0 \ .$$

We use the latter constraint, since it has a nice intuitive interpretation regarding the distribution, namely,

$$\int\limits_{\text{principal recommends } a_2} [R_2 - R_1]d\pi \ .$$

Consider a partition policy that is given by the sets $\{I_t\}$; in this case we have:

$$\begin{aligned}&\Pr(\text{principal recommend } a_2) * E(R_2 - R_1|(\text{principal recommends } a_2)\\ &= \int\limits_{R_1 \in \cup_{\tau<t} I_\tau, R_2 > R_1} [R_2 - R_1]d\pi + \int\limits_{R_1 \in I_t} [\mu_2 - R_1]d\pi \ .\end{aligned}$$

The first integral represents *"exploitation"*, which is defined as the benefit for the agent in the event that the principal is informed about both actions,

i.e., $R_1 \in \cup_{\tau<t} I_\tau$. Obviously this integrand is positive. The second integral, the *"exploration"* part, represents the loss in the case where the principal wishes to explore and agent t is the first agent to try the second action. We will show that in the optimal mechanism this integrand is negative.

Hence, for partition mechanisms we can express the IC constraint as

$$\int\limits_{R_1 \in \cup_{\tau<t} I_\tau, R_2 > R_1} [R_2 - R_1] d\pi + \int\limits_{R_1 \in I_t} [\mu_2 - R_1] d\pi \geq 0 \,. \tag{1}$$

Alternatively, this can be expressed as

$$\int\limits_{R_1 \in \cup_{\tau<t} I_\tau, R_2 > R_1} [R_2 - R_1] d\pi \geq \int\limits_{R_1 \in I_t} [R_1 - \mu_2] d\pi \,.$$

The following lemma shows that it is sufficient to consider the IC of action a_2.

Lemma 3 *Assume that the recommendation $x_t = a_2$ to agent t is IC. Then the recommendation $x_t = a_1$ is also IC.*

Proof: Let $K_t = \{(R_1, R_2) | x_t = a_2\}$ be the event in which the recommendation to agent t is $x_t = a_2$. If $K_t = \emptyset$ then the lemma follows since $E[R_1 - R_2] > 0$. Otherwise $K_t \neq \emptyset$ and because the recommendation $x_t = a_2$ is IC we must have $E[R_2 - R_1 | K_t] \geq 0$. Recall however that by assumption $E[R_2 - R_1] \leq 0$.

Now, since

$$E[R_2 - R_1] = E[R_2 - R_1 | K_t] \Pr[K_t] + E[R_2 - R_1 | \neg K_t] \Pr[\neg K_t] \leq 0,$$

it has to be the case that $E[R_2 - R_1 | \neg K_t] \leq 0$ which in particular implies that recommending $x_t = a_1$ is IC in the case of $\neg K_t$. □

3.4 Optimality of the Threshold Policy

Definition 3 *A threshold policy is a partition policy in which the sets I_t are ordered intervals. Formally, it has a partition of R to intervals$\{I_t\}_{t=2}^{T+1}$ where $I_2 = (-\infty, i_2]$, $I_t = (i_{t-1}, i_t]$ and $I_{T+1} = (i_T, \infty)$.*

Note that I_{T+1} contains all the realizations of R_1 after which the planner recommends action 1 for all agents $t \in \{1, ..., T\}$. We shall associate the $T+1$ interval with a factitious agent. While the definition of a threshold policy is clear, one can define policies that are threshold policies up to measure zero events and achieve the same outcome. This observation is important when we prove optimality of threshold policies. For that purpose it is important to note that if $\{I_j\}_{j=2}^{T+1}$ are not (up to measure zero events) ordered intervals then there exist indexes $t_2 > t_1$ and sets $B_1 \subseteq I_{t_1}$and $B_2 \subset I_{t_2}$ such that: (1) $\sup B_2 < \inf B_1$ and (2) $\Pr[B_1], \Pr[B_2] > 0$.

The following simple claim establishes that in every period, the planner will do as much exploration as the *IC* condition allows.

Lemma 4 *Let Π^* be an optimal partition policy and assume that in Π^* agent $t+1 \geq 3$ explores with some positive probability (i.e., $\Pr[I_{t+1}] > 0$). Then agent t has a tight IC constraint.*

Proof: Assume by way of contradiction that agent t does not have a tight *IC* constraint. Then we can "move"part of the exploration of agent $t+1$ to agent t, and still satisfy the *IC* constraint. The average reward will only increase, since agent $t+1$, rather than exploring, will do the better of the two actions, in the event that agent t explores instead of doing action a_1. To be precise, assume that the *IC* condition for agent t does not hold with equality. That is,

$$\int_{R_1 \in I_t} [R_1 - \mu_2] d\pi \quad < \int_{R_1 \in \cup_{\tau<t} I_\tau, R_2 > R_1} [R_2 - R_1] d\pi \tag{2}$$

Recall that I_t consists of those values r_1 for which agent t is the first to explore action a_2 when $R_1 = r_1$. By assumption we have $\Pr[I_{t+1}] > 0$. Note that the RHS of (2) does not depend on I_t. Therefore, we can find a subset $\hat{I} \subset I_{t+1}$ where $\Pr[\hat{I}] > 0$ and then replace the set I_t with $I'_t = I_t \cup \hat{I}$ and the set I_{t+1} with $I'_{t+1} = I_{t+1} - \hat{I}$ and still keep the IC constraint. The expected average reward increases, since the only difference is when $R_1 \in \hat{I}$ and hence the only change is in the expected rewards of agent t and $t+1$. Before the change, the expected sum of rewards of agents t and $t+1$, conditional on $R_1 \in \hat{I}$, were $\mu_2 + E[R_1|R_1 \in \hat{I}]$, while the new sum of expected rewards (again conditional on $R_1 \in \hat{I},$) is $\mu_2 + E[\max\{R_1, R_2\}|R_1 \in \hat{I}]$, which is strictly larger (since the prior is continuous). The IC constraint of agent $t+1$ still holds, since we only removed exploration. None of the other agents is affected by this modification. Therefore, we reached a contradiction that the policy is optimal. □

Lemma 5 *Assume that policy* Π *is a partition policy and let* B *include the values of the first action which are below the expectation of the second action, and are not in* I_2*. i.e.,*[1]

$$B = \{r_1 : r_1 \leq \mu_2,\ r_1 \notin I_2\}.$$

If $\Pr[B] > 0$ *then a policy* Π' *which is similar to* Π *except that now* $I'_2 = B \cup I_2$ *and* $I'_t = I_t - B$ *for* $t \geq 3$*, is a recommendation policy with a higher expected average reward.*

Proof: Consider the policy Π and let $B_t = B \cap I_t$ for $t \geq 3$. Because Π is a recommendation policy, agent t finds it optimal to follow the recommendations and in particular to use action a_2 when recommended. Next consider the policy Π' and observe that the incentives of agent t to follow the recommendation to use action a_2 are stronger now because for $R_1 \in B_t$

[1]Recall that we assume that $\Pr[R_1 < \mu_2] > 0$.

his payoff in Π is R_2 while in Π' it is $\max\{R_1, R_2\}$. The agents t between 3 and T have a stronger incentive to follow the recommendation, since now in the event of $R_1 \in B_t$ we recommend the better of the two actions rather than a_1. Because $R_1 < \mu_2$ it is immediate that expected average rewards in Π' are higher than in Π.

For agent 2 we have only increased the IC, since $E[R_2 - R_1 | R_1 \in B] \geq 0$. $\square$

The discussion so far allows us to restrict attention to partition policies in which: (i) once both R_1 and R_2 are observed, the policy recommends the better action, (ii) the *IC* constraint is always tight, and (iii) the set $I_2 \supseteq \quad (-\infty, \mu_2]$. Next, we will argue that we should also require the policy to be a threshold policy. Recall that for a non-threshold policy there exist indexes $t_2 > t_1$ and sets $B_1 \subseteq I_{t_1}$and $B_2 \subset I_{t_2}$ such that: (1) $\sup B_2 < \inf B_1$ and (2) $\Pr[B_1], \Pr[B_2] > 0$.

A useful tool in our proof is an operation we call *swap* that changes a policy Π to a policy Π'.

Definition 4 *A swap operation modifies the recommendations of two agents t_1 and $t_2 > t_1$. It takes a partition policy Π and subsets $B_1 \subset I_{t_1}$, $B_2 \subset I_{t_2}$ where $\sup B_2 < \inf B_1$ to constructs a partition policy Π' such that $I'_{t_1} = I_{t_1} \cup B_2 - B_1$ and $I'_{t_2} = I_{t_2} \cup B_1 - B_2$, while other sets are unchanged, i.e., $I'_t = I_t$ for $t \notin \{t_1, t_2\}$. We say that a swap is proper if*

$$\int_{R_1 \in B_1} [\mu_2 - R_1] d\pi \quad = \int_{R_1 \in B_2} [\mu_2 - R_1] d\pi.$$

Since $(-\infty, \mu_2] \subseteq I_2$ we conclude that if the swap operation is proper then for all $R_1 \in B_2 \cup B_1$ we have $R_1 > \mu_2$ which in particular implies that $\Pr[B_2] > \Pr[B_1]$.

Lemma 6 *Let Π be a recommendation policy and let Π' be the policy resulting from a proper swap. Then Π' is a recommendation policy in which the*

expected rewards of all agents are at least as high as in Π *and for some agents they are strictly higher.*

Proof: Since the swap operation is proper we have $\inf B_1 > \sup B_2$, $\Pr[B_2] > \Pr[B_1]$ and

$$\int\limits_{R_1 \in B_1} [\mu_2 - R_1] d\pi \quad = \int\limits_{R_1 \in B_2} [\mu_2 - R_1] d\pi.$$

First we show that the swap does not change the expected reward of agent t_1 conditional on a recommendation to choose action a_2. From the perspective of agent t_1, the change is that in the case where $r_1 \in B_1$ the action recommended to him at Π' is a_1 rather than the action a_2 which is recommended to him at Π, and in the case where $r_1 \in B_2$ it is a_2 (at Π') rather than a_1 (at Π). Since the swap operation is proper, his IC constraint at Π' can be written as:

$$\int\limits_{R_1 \in \cup_{\tau<t_1} I_\tau, R_2 > R_1} [R_2 - R_1] d\pi + \int\limits_{R_1 \in I_{t_1}} [\mu_2 - R_1] d\pi + \int\limits_{R_1 \in B_2} [\mu_2 - R_1] d\pi - \int\limits_{R_1 \in B_1} [\mu_2 - R_1] d\pi$$
$$= \int\limits_{R_1 \in \cup_{\tau<t_1} I_\tau, R_2 > R_1} [R_2 - R_1] d\pi + \int\limits_{R_1 \in I_{t_1}} [\mu_2 - R_1] d\pi \geq 0.$$

Therefore the swap does not change the expected reward of agent t_1 and Π' satisfies IC for this agent.

Next consider all agents *except agents* t_2 and t_1. Observe first that all agents $t < t_1$ and $t > t_2$ do not observe any change in their incentives (and rewards) and we are left with agents t where $t_1 < t < t_2$. The expected rewards of these agents can only increase because the effect of the swap is only on the first integral $\int\limits_{R_1 \in \cup_{\tau<t} I_\tau, R_2 > R_1} [R_2 - R_1] d\pi$ of the IC constraint (see Eq (1)) which increases as a result of the swap because instead of the set $\cup_{\tau<t} I_\tau$ we now have $\cup_{\tau<t} I_\tau \cup B_2 - B_1$ and $\sup B_2 < \inf B_1$.

Thus, it is left for us to analyze the incentives and rewards of agent t_2 (and

only when $t_2 \leq T$) to follow the recommendation to choose action a_2. First observe that if $r_1 \notin B_1 \cup B_2$ then Π and Π' are identical, and hence the only case to consider is when $r_1 \in B_1 \cup B_2$. The expected reward under Π conditional on $r_1 \in B_1 \cup B_2$ is

$$\frac{1}{\Pr[B_1 \cup B_2]} \left[\int_{R_1 \in B_1, R_2 > R_1} [R_2 - R_1] d\pi + \int_{R_1 \in B_2} [\mu_2 - R_1] d\pi \right],$$

and the expected reward under Π' is

$$\frac{1}{\Pr[B_1 \cup B_2]} \left[\int_{R_1 \in B_2, R_2 > R_1} [R_2 - R_1] d\pi + \int_{R_1 \in B_1} [\mu_2 - R_1] d\pi \right],$$

We would like to show that

$$\int_{R_1 \in B_1, R_2 > R_1} [R_2 - R_1] d\pi + \int_{R_1 \in B_2} [\mu_2 - R_1] d\pi \quad < \int_{R_1 \in B_2, R_2 > R_1} [R_2 - R_1] d\pi + \int_{R_1 \in B_1} [\mu_2 - R_1] d\pi,$$

which is equivalent to showing that (recall that the swap is proper)

$$\int_{R_1 \in B_1} \int_{R_2 > R_1} [R_2 - R_1] d\pi \quad < \int_{R_1 \in B_2} \int_{R_2 > R_1} [R_2 - R_1] d\pi.$$

The last inequality is a simple consequence of

$$\Pr[B_2] > \Pr[B_1] \textit{ and } \inf B_1 > \sup B_2.$$

This again implies that the *IC* constraint is satisfied for this agent and that the swap operation increases his rewards. □

Lemma 6 implies that an optimal policy must be a threshold policy. That is, the sets $\{I_t\}_{t \in T}$ are restricted to being a set of intervals. Moreover, the *IC* constraint holds for any agent t provided that there is a positive probability

that agent $t+1$ will be asked to explore.

Thus, to fully characterize the optimal policy, all that is left for us to identify is the threshold θ, such that if $r_1 \leq \theta$ then some agent t will explore action a_2. The threshold θ, which is a function of T, together with the intervals $\{I_t\}_{t=2}^{T}$, fully characterize the optimal policy Π^{opt}. Solving for θ is the topic of the following subsection.

3.5 The Optimal Threshold Policy

Consider first the case where T is infinite. In this case exploration is maximized as the planner wishes to explore for any realized value of the first action, r_1. The optimal policy is defined by an increasing sequence of thresholds $i_{1,\infty} < i_{2,\infty}...$ where for $t = 2$

$$\int_{R_1=-\infty}^{i_{2,\infty}} [R_1 - \mu_2] d\pi = 0$$

For $t > 2$ as long as $i_{t,\infty} < \infty$, we have

$$i_{t+1,\infty} = \sup \left\{ i | \int_{R_1 \leq i_t, R_2 > R_1} [R_2 - R_1] d\pi \geq \int_{R_1=i_{t,\infty}}^{i} [R_1 - \mu_2] d\pi \right\}$$

If $i_{t,\infty} = \infty$ then we define $i_{t',\infty} = \infty$ for all $t' \geq t$. Note that if $i_{t+1,\infty} < \infty$ then the above supremum can be replaced with the following equality

$$\int_{R_1 \leq i_t, R_2 > R_1} [R_2 - R_1] d\pi = \int_{R_1=i_{t,\infty}}^{i_{t+1,\infty}} [R_1 - \mu_2] d\pi \tag{3}$$

Consider the case when T is finite. As we shall see the planner will ask fewer agents to explore. Consider the t-th agent. The RHS is the expected loss due to exploration by the current agent is $r_1 - \mu_2$. The expected gain in

exploitation, if we explore, is $(T-t)E[\max\{R_2 - r_1, 0\}]$. We will set the threshold θ_t for agent t to be the maximum r_1 for which it is beneficial to explore. Let θ_t be the solution to

$$(T-t)E[\max\{R_2 - \theta_t, 0\}] = \theta_t - \mu_2$$

If there are $T-t$ agents left then θ_t is the highest value for which it is still optimal to explore; Note that θ_t is decreasing in t. Our main result is:

Theorem 7 *The optimal policy,* Π^{opt}*, is defined by the sequence of thresholds*

$$i_{t,T} = \min\{i_{t,\infty}, \theta_\tau\},$$

where τ *is the minimal index for which* $i_{t,\infty} > \theta_t$*.*

Next, we argue that even when T is arbitrarily high, exploration is limited to a finite number of agents.

Theorem 8 *Let* $t^* = \min\{t | i_t = \infty\}$*; then* $t^* \leq \frac{\mu_1 - \mu_2}{\alpha}$ *where*

$$\begin{aligned}\alpha &= \int_{R_1 \leq i_2, R_2 > R_1} [R_2 - R_1] d\pi \\ &\geq \Pr[R_2 \geq \mu_2] \cdot \Pr[R_1 < \mu_2] \cdot (E[R_2 | R_2 \geq \mu_2] - E[R_1 | R_1 < \mu_1])\end{aligned}$$

Since t^* *is finite, the principal is able to explore both actions after* t^* *agents.*

The proof appears in the Appendix but we can provide the intuition here. Consider (3), the LHS represents the gain agent t expects to receive by following the recommendation of the principal who has already tested both alternatives. It is an increasing sequence as for higher t the planner becomes better informed. This implies that this terms can be bounded from bellow when we consider agent $t = 2$. The RHS represents the expected loss the agent expects to experience when he is the first agent to try the second

alternative. The sum of the RHS over all t is the difference in means $\mu_1 - \mu_2$. The proof is based on these two observations when we sum the LHS and RHS.

The above theorem has important implications. Consider the first-best outcome in which the principal can force agents to choose an action. The above theorem implies that for any T the aggregate loss of the optimal mechanism as compared to the first-best outcome is bounded by $\frac{(\mu_1-\mu_2)^2}{\alpha}$. As a result we conclude that:

Corollary 9 *As T goes to infinity the average loss per agent as compared to the first-best outcome converges to zero at a rate of $1/T$. Apart from a finite number of agents, t^*, all other agents are guaranteed to follow the optimal action.*

The above theorem has important implications. Consider the first-best outcome in which the principal can force agents to choose an action. The above theorem implies that for any T the aggregate loss of the optimal mechanism as compared to the first-best outcome is bounded by $\frac{(\mu_1-\mu_2)^2}{\alpha}$.

4 Cash Incentives- A planner with a budget

We first suppose that the planner has a budget and can spend up to $\$X$ to induce exploration and improve total welfare. We then endogenize the choice of X. We assume that agents' utility is additive with respect to cash incentives. If the principal offers a cash incentive of x to follow his recommendation to take the second action then the IC constraint is given by:

$$E(R_2 - R_1|\text{principal recommends } a_2) \geq -x$$

We assume that the principal maximizes social welfare subject to the budget constraint. One could also solve for the optimal X by comparing it to the

total benefit from being able to induce more exploration. Given that our result holds for an arbitrary X this would not change our conclusion so we focus on the former formulation. We also assume that the offer is binding, namely, the cash incentive is dependent on performing the recommended action.

Assuming X is exogenously fixed, then the principal follows a disclosure policy similar to the one described in Section 3 when transfers were not allowed with two small differences. One is that for some distributions it might be optimal to explore first action two even though $\mu_1 > \mu_2$. In what follow we assume that this is not the case and extending the result to follow to this case is immediate. The more substantial difference is that now the planner can induce more exploration by promising cash payments. The proof follows similar logic to the proofs above and consequently, we will do here with an outline only and omit some of the details.

If $X \geq \mu_1 - \mu_2$ the principal can convince the second agent to explore regardless of the realization of R_1 and he can obtain the efficient outcome (ignoring of course the cost of X).[2] In the more interesting case when $X < \mu_1 - \mu_2$, the principal will use X to convince the second agent to explore only if R_1 is not too high and no cash incentive is offered to agent $t > 2$. In particular

$$i_2 = \sup \left\{ i | \quad X \geq \int_{-\infty}^{i} [R_1 - \mu_2] d\pi \right\}.$$

To see why this must be the case note first that as Lemma 4 can be generalized along the following lines; If agent $t+1 \geq 3$ explores with some positive probability (i.e., $\Pr[I_{t+1}] > 0$) then agent t has a tight IC constraint. Furthermore, using the same 'swap' argument one can also show that the

[2]Note that the planner can use even slightly smaller X to get this effect, since there is some value θ_T, such that if $R_1 > \theta_T$ the optimal policy never explores. Hence conditioned on exploring, the value is slightly less than μ_1.

optimal disclosure policy is threshold policy, $\{I_t\}_{t=2}^{T+1}$ where $I_2 = (-\infty, i_2]$, $I_t = (i_{t-1}, i_t]$ and $I_{T+1} = (i_T, \infty)$. Thus the principal's optimal policy Π is given by the intervals $\{I_t\}_{t=2}^{T+1}$ and a sequence of cash incentives $\{x_t\}$ where $\sum x_t \leq X$.

Assume by contradiction that agent $t+1 > 2$ is provided cash incentive, $x_{t+1} > 0$ to use the second alternative. Consider a policy Π^* which is the same as Π except that now the transfers are

$$x_t^* = x_t + x_{t+1} \quad \text{and} \quad x_{t+1}^* = 0$$

and we adjust the intervals accordingly:

$$I_t^* = (i_{t-1}, i^*], \quad \text{and} \quad I_{t+1}^* = (i^*, i_{t+1}]$$

where i^* is defined by:

$$\int_{R_1=i_{t-1}}^{i^*} [\mu_2 - R_1] d\pi = -x_t^*$$

Since,

$$\int_{R_1=i_{t-1}}^{i_{t+1}} [\mu_2 - R_1] d\pi = \int_{R_1=i_{t-1}}^{i_t} [\mu_2 - R_1] d\pi + \int_{R_1=i_t}^{i_{t+1}} [\mu_2 - R_1] d\pi = x_t + x_{t+1} = x_t^*,$$

we have that

$$\int_{R_1=i^*}^{i_{t+1}} [\mu_2 - R_1] d\pi = 0,$$

and it follows that the IC constraint is satisfied for agents t and $t+1$. Finally, we reach a contradiction as exploration is expedited, and hence we improve the expected average return. Since we can apply this argument to any agent $t \geq 3$, we have established the following theorem,

Theorem 10 *In the optimal policy the monitory incentives would be given only to agent* 2.

We first describe the optimal policy for T being infinite, we set i_2 so that:

$$\int_{R_1=-\infty}^{i_{2,\infty}} [R_1 - \mu_2] d\pi = -X$$

For $t > 2$ and as long as $i_{t,\infty} < \infty$, we have:

$$i_{t+1,\infty} = \sup \left\{ i | \int_{R_1 \leq i_{t,\infty}, R_2 > R_1} [R_2 - R_1] d\pi \geq \int_{R_1=i_{t,\infty}}^{i} [R_1 - \mu_2] d\pi \right\}$$

As in the case with no budget when one considers a finite T then one needs to adjust the above intervals. At time t the planner compares the potential benefits to the remaining $T-t-1$ agents to the cost of t taking a suboptimal action. This trade-off is given by θ_t and one then defines the intervals as $i_{t,T} = \min\{i_{t,\infty}, \theta_\tau\}$.

Theorem 11 *The optimal policy with cash incentive budget* X, Π_X^{opt}, *is defined by the sequence of thresholds*

$$i_{t,T} = \min\{i_{t,\infty}, \theta_\tau\},$$

where τ *is the minimal index for which* $i_{t,\infty} > \theta_t$, *and agent* 2 *is offered a cash incentive of* $\min\{\mu_1 - \mu_2, X\}$.

Consider now the case of a planner who raises X at a cost of δX for some $\delta > 0$.[3] Absent such cost, X, can be viewed as a transfer among agents that has no welfare implication; the social planner simply raises X through taxes

[3]One can imagine that the social planner obtains the needed money by collecting taxes and the δ represents the distortion caused by this.

and uses it to subsidize exploration. The cost can be viewed as a deadweight loss that is associated with taxation and does affect aggregate welfare. The above result implies that when X is not too high then the social planner will choose to conceal some information. We argue that when X is chosen endogenously for $\delta > 0$ that is arbitrary small, the social planner will choose X not too high so that he will indeed hide some information. The reason why this is true is that the benefit from asking the second agent to explore is decreasing in X to zero.

Theorem 12 *The optimal X when the cost is δX is X^*such that $\Pi^{opt}_{X^*}$ satisfies $i_{2.T} < \infty$.*

References

[1] Bikhchandani S., D. Hirshleifer and I. Welch (1992): "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades", *Journal of Political Economic* Vol. 100, No. 5.

[2] Bolton P. and C. Harris (1999): "Strategic Experimentation", *Econometrica*, 67, 349–374.

[3] Gershkov A. and B. Szentes (2009): "Optimal Voting Schemes with Costly Information Acquisition" *Journal of Economic Theory*, Vol.144 (1), 36-68.

[4] Gustavo M. (2012): "Motivating Innovation." *Journal of Finance* (forthcoming).

[5] Kamenica E, and M. Gentzkow (2011) "Bayesian Persuasion." *American Economic Review,* 101, 2590-2615.

[6] Keller G., S. Rady and M.W. Cripps (2005): "Strategic Experimentation with Exponential Bandits" , Econometrica, vol. 73, 39-68.

[7] Martimort, D. and S. Aggey (2006): "Continuity in mechanism design without transfers," Economics Letters, vol. 93(2), 182-189.

[8] Rothschild M. (1974): "A Two-Armed Bandit Theory of Market Pricing", Journal of Economic Theory, 9, 185–202.

A Missing proofs

Proof of Theorem 8: Given our characterization it is sufficient to focus on the case when $T = \infty$. Consider the summation oh the RHS in (3):

$$\sum_{t=2}^{\infty} \int_{R_1=i_{t,\infty}}^{i_{t+1,\infty}} [R_1 - \mu_2] d\pi = \lim_{t\to\infty} \int_{R_1=i_{2,\infty}}^{i_{t,\infty}} [R_1 - \mu_2] d\pi$$

since $\int_{R_1=-\infty}^{i_{2,\infty}} [R_1 - \mu_2] d\pi = 0$ and since $\int_{R_1\leq x} [R_1 - \mu_2] d\pi$ is increasing in x we conclude that:

$$\sum_{t=2}^{\infty} \int_{R_1=i_{t,\infty}}^{i_{t+1,\infty}} [R_1 - \mu_2] d\pi \leq \lim_{x\to\infty} \int_{R_1\leq x} [R_1 - \mu_2] d\pi = \mu_1 - \mu_2$$

Looking at the summation of the LHS

$$\sum_{t=2}^{\infty} \int_{R_1\leq i_t, R_2>R_1} [R_2 - R_1] d\pi$$

we note that $\int_{R_1\leq x, R_2>R_1} [R_2 - R_1] d\pi$ is increasing in x. The fact that i_t is increasing in t implies that if we let

$$\alpha \equiv \int_{R_1\leq i_2, R_2>R_1} [R_2 - R_1] d\pi$$

we then have

$$\alpha \leq \int\limits_{R_1 \leq i_t, R_2 > R_1} [R_2 - R_1] d\pi$$

Hence, this sum can be bounded from below by $t^* \alpha$, which implies the claim. □