# The Communication Cost of Selfishness\*

Ronald Fadel<sup>†</sup>

Ilya Segal<sup>‡</sup>

January 26, 2006

#### Abstract

We consider the amount of communication required to implement a given decision rule when the mechanism must be ex post or Bayesian incentive compatible. In ex post incentive compatibility, the communication protocol must reveal enough information to calculate monetary transfers to the agents to motivate them to be honest (agents' payoffs are assumed to be quasilinear in such transfers). For Bayesian incentive compatibility, the protocol may need to hide some information from the agents to prevent deviations contingent on the information. In both cases, the selfishness of the agents can substancially increase the communication costs. We provide an exponential upper bound on the communication cost of selfishness, which is tight in the Bayesian setting. Whether this exponential upper bound is ever achieved in the ex post setting remains an open question. We examine some extensions of our initial setting. In particular we show that for the average-case communication complexity measure, the communication cost of selfishness may be arbitrarily large in both ex post and Bayesian settings. We also examine some special cases in which the communication cost of selfishness proves to be very low, in particular when we want to implement efficiency.

## 1 Introduction

This paper straddles two literatures on allocation mechanisms. One literature, known as "mechanism design," examines the agents' incentives in the mechanism. Appealing to the "revelation principle," the literature focuses on "direct revelation mechanisms" in which agents fully describe their preferences, and checks their incentives to do so truthfully (e.g., [Mas-Colell et al., 1995, Chapter 23]). However, full revelation of private information would be prohibitively costly in most practical settings. For example, in a combinatorial auction with L objects, full revelation would require describing a utility value of each of the  $2^L - 1$  nonempty bundles of objects, which with L = 30 would take more than 1 billion numbers. The other literature examines how much communication, measured with the number of bits or real variables, is required in order to compute the social outcome, assuming that agents communicate truthfully (e.g., [Kushilevitz and Nisan, 1997], [Nisan and Segal, 2004], [Segal, 2005], and references therein). However, in most practical settings we should expect agents to communicate strategically to maximize their own benefit.

This paper considers how much communication is required in order to implement a given decision rule when agents are selfish. The mechanism designer can use two instruments to induce agents to report truthfully equilibrium: First, along with the allocation she could use the communication

<sup>\*</sup>The authors aknowledge the support of the National Science Foundation (grants SES 0318447, ITR 0427770). Preliminary results from this work were reported in an extended abstract (Fadel and Segal 2005).

<sup>&</sup>lt;sup>†</sup>Department of Economics, Stanford University, Stanford CA 94305, email rfadel@cs.stanford.edu.

<sup>&</sup>lt;sup>‡</sup>Department of Economics, Stanford University, Stanford CA 94305, email ilya.segal@stanford.edu.

protocol to compute transfers to the agents (we assume that agents' payoffs are quasi-linear in such transfers). Second, in the course of computing the outcome, the designer may hide some information from the agents (i.e., create information sets), thus reducing the set of contingent deviations available to them. Both the need to compute motivating transfers and the need to hide information from agents may increase the communication cost relative to that of computing the allocation when the agents are honest. Indeed, we offer simple examples where a protocol computing the desired decision rule cannot be incentivized using transfers or information sets (even though the same rule can be implemented in a direct revelation mechanism), and we have a strictly positive "communication cost of selfishness" (for short, overhead) — the additional communication cost required when agents report selfishly rather than truthfully.

The focus of the paper is on the case in which agents have private and independently drawn valuations over outcomes, and the communication cost is defined as the maximal number of bits sent during the execution of the mechanism. We consider implementation in two distinct equilibrium concepts: Ex Post Incentive Compatibility (EPIC), in which an agent should report honestly even if he somehow finds out other agents' private information, and Bayesian-Nash Incentive Compatibility (BIC), in which an agent reports honestly given his beliefs about other agents' information. In each case, we show that the communication cost of selfishness may be strictly positive. However, the reasons for the overhead differ between the two cases: In the EPIC case, there is no need to hide information from agents, and the overhead comes entirely from the need to compute motivating transfers. In the BIC case, in contrast, computing transfers is not a problem, and the overhead comes from the need to hide information from the agents, to eliminate some of their contingent deviations.

We begin our analysis by showing that, both for EPIC and BIC implementation, any simultaneous-communication protocol is incentivizable with some transfers. Intuitively, in such a protocol, we fully observe all agents' strategies, and this proves sufficient to compute incentivizing transfers. The problem with sequential (extensive-form) communication protocols is that agents have contingent strategies that are not fully revealed in the course of the protocol. Yet, such sequential communication is typically needed to minimize the communication cost.

Next we observe that starting with any sequential protocol, we can convert it into a simultaneous protocol computing the same decision rule — the "normal form" game in which agents announce their complete contingent strategies in the original protocol. The simultaneous protocol can then be incentivized. If the communication cost of the original sequential protocol was b bits, then the number of bits in the normal-form game will be at most  $2^b - 1$ . This gives an upper bound on the communication cost of selfishness. However, the bound is exponential, and we would like to know whether the bound is ever achieved.

For EPIC implementation, we do not know whether the exponential upper bound is ever achieved – in fact, we do not have any examples where the overhead is large. On the other hand, for the BIC case, we do have an example with an exponential overhead. The example can be interpreted as having an expert with private knowledge and a private utility function, and a manager with a private goal for how the expert's knowledge should be used. The expert will reveal his knowledge truthfully if he does not know how the manager plans to use it, but this revelation will take exponential communication in the number of outcomes. Communication cost would be exponentially lower if the manager first announces his goals and then the expert says how to achieve it, but this communication would not be incentive-compatible — the expert would manipulate the outcome to her advantage. We show that any communication in which the expert's incentives are satisfied would be exponential in the number of outcomes – almost as long as full revelation of the expert's knowledge.

This example notwithstanding, we do find several special cases in which the communication cost of selfishness is low. In particular, when the decision rule to be implemented maximizes the total surplus (the sum of the agents' valuations)., the communication cost of selfishness proves to be low, both in the EPIC and BIC case.

Finally, we consider several extensions of the model. First, we consider the communication cost measured as the *average-case* rather than worst-case number of bits sent, given a probability distribution over agents' information. We show that the average-case communication cost of selfishness could be unbounded, at least for some probability distributions. Next, we show that the communication cost of selfishness could be unbounded when agents' valuations are correlated rather than independent (for BIC implementation), or interdependent rather than private (for EPIC implementation).

## 2 Related Literature

A number of papers have proposed incentive-compatible indirect communication mechanisms in various special settings. The first paper we know of is Reichelstein [Reichelstein, 1984], who considered incentive compatibility in nondeterministic real-valued mechanisms, and showed that the communication cost of selfishness in achieving efficiency is low. Lahaie and Parkes [Lahaie and Parkes, 2004] characterized the communication problem of finding Vickrey-Groves-Clarke (VCG) transfers as that of finding a "universal price equilibrium," but did not examine the communication complexity of finding such an equilibrium, or the possibility of implementing efficiency using non-VCG transfers. Neither paper examined the communication complexity of decision rules other than surplus maximization. For an analysis of the communication requirements of incentive-compatible mechanisms in networks, see Feigenbaum et al. [Feigenbaum et al., 2002].

A few papers on incentive-compatible communication have considered a "dual" question: instead of asking how much communication is needed to achieve a given goal, it asks how to maximize a given objective function subject to a fixed communication constraint. In one literature, the objective is to maximize the profits of one of the agents subject to other agents' participation constraints. See, e.g., [Green and Laffont, 1987],[Melumad et al., 1992], and a recent survey by [Mookherjee, 2006]. A similar question is studied by [Johari, 2004], who instead focuses on the efficiency objective.

Finally, the literature on communication without commitment ("cheap talk") has offered examples in which incentive-compatible communication would require a large number of stages (e.g., [Forges, 1990]). In contrast, our mechanism commits to an outcome as a function of messages, yet we find the communication cost as measured in bits to be potentially high. (But it would be possible to send all the bits in one stage – e.g., in a direct revelation mechanism).

# 3 Communication With Honest Agents: Communication Complexity

The concept of communication complexity, introduced by Yao [Yao, 1979] and surveyed in [Kushilevitz and Nisan, 1997], describes how much communication is needed for agents from a set  $I = \{1, ..., I\}$  to compute the value of a function  $f: \prod_{i \in I} U_i \to X$  when, for every input  $(u_1, ..., u_I)$ , each agent  $i \in I$  knows privately only  $u_i \in U_i$ , which we refer to as agent i's "type". Communication is modeled using the notion of a protocol. In the language of game theory, a protocol is

simply an extensive-form game along with the agents' strategies in it. Without loss of generality, the communication complexity literature restricts attention to games of perfect information (i.e., each agent observes the history of the game). Also, we restrict attention to protocols in which each agent has two possible moves (messages, interpreted as sending a bit) at a decision node, since any message from a finite set can be coded using a fixed number of bits. Formally,

**Definition 1.** A protocol  $\mathcal{P}$  with agents  $I = \{1, ..., I\}$  over state space  $U = \prod_{i \in I} U_i$  and outcome space X is a binary tree, with set of nodes N and set of leaves  $L \subset N$ , where:

- The set  $N \setminus L$  of non-leaf nodes (i.e., decision nodes) is partitioned into I subsets  $N_1, \ldots, N_I$ , with  $N_i$  representing the set of decision nodes of agent  $i \in I$ .
- Each leaf  $l \in L$  of the tree is labeled with an outcome  $x(l) \in X$ .
- Each agent  $i \in I$  has a strategy plan  $\sigma_i : U_i \to \{0,1\}^{N_i}$ , where  $\{0,1\}^{N_i}$  is the set of the agent's possible strategies in the protocol lists of moves made at his decision nodes.<sup>1</sup>

For each strategy profile  $s = (s_1, \ldots s_I) \in \prod_{i \in I} \{0, 1\}^{N_i}$ , let  $g(s) \in L$  denote the leaf l that is reached when each agent i follows the strategy  $s_i$ . The function  $f: U \to X$  computed by protocol  $\mathcal{P}$ , which is denoted by  $Fun(\mathcal{P})$ , is defined by  $f = x \circ g \circ \sigma$ .

Given a protocol  $\mathcal{P}$ , it is convenient to define for each node  $n \in N$  its "legal domain"  $U(n) \subset U$  as the set of inputs on which node n is reached. For example, for the root r of the tree, U(r) = U. By forward induction on the tree, it is easy to see that the legal domain at each node n is a product set  $U(n) = \prod_{i \in I} U_i(n)$ , using the fact that each agent's prescribed move at any node depends only on his own type. Without loss of generality, we consider only protocols such that all the nodes n have a non-empty legal domain U(n).

The depth  $d(\mathcal{P})$  of a protocol  $\mathcal{P}$  is the maximum number of edges between the root and a leaf — i.e., the number of bits sent in the protocol in the worst case.<sup>2</sup>

**Definition 2.** For each function  $f: U \to X$ , we define CC(f), the communication complexity of f, as the depth of the shallowest protocol  $\mathcal{P}$  that computes f, i.e.,  $CC(f) = \min_{\mathcal{P}: Fun(\mathcal{P}) = f} d(\mathcal{P})$ .

Hence the communication complexity CC(f) of a function f is the minimal communication cost required to compute the function, while the communication cost of a protocol is the maximal number of bits sent during the execution of the protocol.

# 4 Communication With Selfish Agents: Binary Dynamic Mechanisms

#### 4.1 The Formalism

In our case, the function to be computed is the decision rule to be implemented. The protocol may also compute transfers to the agents. We now assume that each agent has preferences described by

<sup>&</sup>lt;sup>1</sup>It is customary in game theory to call the "strategy" of agent i the whole function  $\sigma_i$ , since the agent's type  $u_i \in U_i$  can be interpreted as a "move of nature" on which his strategy could be contingent. However, for our purposes it is convenient to reserve the term "strategy" to denote the agent's behavior  $s_i \in \{0,1\}^{N_i}$  in the protocol.

<sup>&</sup>lt;sup>2</sup>We consider average-case communication costs in Section 8.1.

his type. With a slight abuse in notations, we will denote also as  $u_i$  the utility function of agent i having type  $u_i$ .<sup>3</sup> The utility function  $u_i: X \to \mathbb{R}$  gives the utility of agent i for each outcome. We assume the utilities are quasi-linear, i.e., the total payoff of agent i having type  $u_i \in U_i$  with outcome  $x \in X$  and transfer  $t_i$  is  $u_i(x) + t_i$ . With quasi-linear utilities, a decision rule f is efficient if it satisfies  $f(u) \in \arg\max_{x \in X} \sum_{i \in I} u_i(x), \forall u \in U$ .

Note that our formalism implicitly assumes the *private valuations* setting, i.e., the preferences of an agent depend on the agent's type only. Relaxation of this assumption will be discussed in Section 8.4.

A protocol induces an extensive-form game, and a strategy in this game for each agent. When agents are selfish, we need to consider their incentives to deviate to other strategies in the game<sup>4</sup>. For every agent i having type  $u_i \in U_i$ , his incentive to follow the prescribed strategy  $\sigma_i(u_i)$  depends on the monetary transfer that the protocol assigns to him along with the outcome. In the Bayesian case, it also depends on how much the agent knows about the other agents' types. We formalize this with the notion of a binary dynamic mechanism:

**Definition 3.** A binary dynamic mechanism (BDM) is a triple  $\langle \mathcal{P}, H, t \rangle$  that satisfies:

- $\mathcal{P}$  is a protocol with set of leaves L and set of decision nodes  $\bigcup_i N_i = N \setminus L$ .
- $H = \bigcup_i H_i$  where each  $H_i$  is a partition of  $N_i$  into information sets<sup>5</sup> satisfying perfect recall<sup>6</sup>. The strategy space  $S_i \subset \{0,1\}^{N_i}$  of each agent i is the set of agent i's possible strategies given his information sets, i.e., all the  $s \in \{0,1\}^{N_i}$  satisfying  $\forall h \in H_i \ \forall n, n' \in h : s(n) = s(n')$ .
- $t = (t_1, \ldots, t_I)$  is a list of transfer functions with  $t_i : L \to \mathbb{R}$ .
- The strategy plan  $\sigma$  of protocol  $\mathcal{P}$  is consistent with the information sets: For every agent i and type  $u_i \in U_i$ ,  $\sigma_i(u_i) \in S_i$ .

When leaf  $l \in L$  is reached, outcome x(l) is implemented and each agent  $i \in I$  receives transfer  $t_i(l)$ .

Note that, as the information sets of agent i become larger (and hence their number Card  $H_i$  is reduced), his strategy space shrinks. We define the communication cost (or just depth)  $d(\mathcal{B})$  of any BDM  $\mathcal{B} = \langle \mathcal{P}, H, t \rangle$  as the depth  $d(\mathcal{P})$  of the protocol  $\mathcal{P}$ .

#### 4.2 Two Particular Classes of BDMs

When considering BDMs, it is useful to define two extreme cases of information revelation to the agents. On the one hand, in some BDMs, agents observe the complete history of messages. Formally, a *Perfect information BDM* (PBDM) is a BDM  $\langle \mathcal{P}, H, t \rangle$  such that every information set  $h \in H$  is a singleton. On the other hand, in some BDMs, the agents never learn anything as

<sup>&</sup>lt;sup>3</sup>Note that this notation still allows two different types to correspond to the same utility function by adding a fictitious outcome that gives a different "payoff" for each of the types.

<sup>&</sup>lt;sup>4</sup>Our restriction to agents following the prescribed strategies is without loss of generality, for we can prescribe any possible strategy profile of the game.

<sup>&</sup>lt;sup>5</sup>This setting can be extended to the case where types are revealed to agents in real time as the mechanism is executed, as long as each agent has enough information at each node to compute the prescribed move, and as long as the knowledge of each agent i at information set  $h \in H_i$  depends only on h and his type  $u_i \in U_i$ .

<sup>&</sup>lt;sup>6</sup>The information sets  $H_i$  of agent i satisfy perfect recall if, for every information set  $h \in H_i$  and every two nodes  $n, n' \in h$ , n and n' have the same history of moves and information sets for agent i. See [Fudenberg and Tirole, 1991, p.81] for more details.

the game unfolds. In this case, message ordering is irrelevant, and we can consider that they all communicate simultaneously as each agent can only send some unconditional message about his type. We call *Simultaneous communication BDM* (SBDM) is a BDM  $\langle \mathcal{P}, H, t \rangle$  in which H reflects the simultaneous communication setting: formally, each agent observes only his own previous moves: two decision nodes n and n' of an agent i will be in the same information set  $h_i \in H_i$  if and only if they correspond to the same history of moves for agent i.

In particular a direct revelation BDM is an SBDM where, for each agent i,  $\sigma_i : U_i \to S_i$  is one-to-one (note that this is only possible with finite type spaces). At the end of the execution of a direct revelation BDM, the state is completely known but no agent has observed anything besides his own moves.

Note that some protocols cannot be used in an SBDM, e.g., their strategy plan may not be compatible with simultaneous communication. A *simultaneous communication protocols* is defined as a protocol that can be part of an SBDM.

**Lemma 1.** For any protocol  $\mathcal{P}$ , there exists a simultaneous communication protocol  $\mathcal{P}'$  such that  $Fun(\mathcal{P}) = Fun(\mathcal{P}')$  and  $d(\mathcal{P}') \le 2^{d(\mathcal{P})} - 1$ .

Proof. Given  $\mathcal{P}$ , construct protocol  $\mathcal{P}'$  by having each agent output his complete strategy in the protocol. The order in which agents output their strategies is irrelevant, and each agent need not consider the other agents' moves, as the moves of each agent in  $\mathcal{P}'$  do not depend on the other agents' behaviors. Hence the protocol is a simultaneous communication protocol, and each run will fully reveal the strategy of all agents. Note that since we require that all nodes have a non-empty domain, the encoding of the strategies in  $\mathcal{P}'$  must be such that all the strategies available to each agent in the protocol correspond to a prescribed strategy in  $\mathcal{P}$  for some type. Finally, by counting the number of possible strategies in  $\mathcal{P}$ , each agent i will outure at most one bit for each of his decision nodes  $n \in N_i$  in  $\mathcal{P}$ , so that the depth of  $\mathcal{P}'$  will be at most  $\sum_{i \in I} \operatorname{Card}(N_i) = \operatorname{Card}(N \setminus L) \leq 2^{d(\mathcal{P})} - 1$ .  $\square$ 

# 4.3 Ex Post Incentive-Compatibility

The concept of ex post incentive-compatibility means that for each input, the agents' strategies constitute a Nash equilibrium even if they know the complete input (i.e., each other's types).

**Definition 4.** BDM  $\langle \mathcal{P}, H, t \rangle$  is Ex Post Incentive Compatible (EPIC) if in any state  $u \in U$ , the strategy profile  $s = (\sigma_1(u_1), \ldots, \sigma_I(u_I)) \in \prod_{i \in I} S_i$  is an ex post Nash equilibrium of the induced game, i.e.,

$$\forall i \in I, \forall s_i' \in S_i: \ u_i(x(g(s))) + t_i(g(s)) \ge u_i(x(g(s_i', s_{-i}))) + t_i(g(s_i', s_{-i})).$$

We say say that the BDM is  $\varepsilon$ -EPIC for some  $\varepsilon > 0$  if in any state  $u \in U$ , the strategy profile is an  $\varepsilon$ -Nash equilibrium of the game, i.e., the above inequalities are violated by at most  $\varepsilon$ .

In words, for every state  $u \in U$ ,  $\sigma_i(u_i)$  is an optimal strategy for agent i whatever the types of the other agents are, as long as they follow their prescribed strategies.<sup>7</sup> When the BDM is EPIC, we say that it implements  $Fun(\mathcal{P})$  in EPIC. It turns out that to check if a BDM is EPIC, we only need to consider the transfer function and the legal domains of the leaves. Formally:

<sup>&</sup>lt;sup>7</sup>In our setting of private valuations, in which agent i's utility does not depend on others' types  $u_{-i}$ , this is equivalent to requiring that agent i's strategy be optimal for him assuming that each agent  $j \neq i$  follows a strategy prescribed for some type  $u_j \in U_j$ . Note that we do not require the stronger property of Dominant strategy Incentive Compatibility (DIC), which would allow agent i to expect agents  $j \neq i$  to use contingent strategies  $s_j \in S_j$  that are inconsistent with any type  $u_j$ , and which would be violated in even the simplest dynamic mechanisms. We discuss dominant strategy implementation in Section 8.3

**Lemma 2.** A BDM  $\langle \mathcal{P}, H, t \rangle$  is EPIC if and only if for every agent  $i \in I$  and every two leaves  $l, l' \in L$ :

$$U_{-i}(l) \cap U_{-i}(l') \neq \emptyset \implies \forall u_i \in U_i(l) : u_i(x(l)) + t_i(l) \ge u_i(x(l')) + t_i(l')$$
 (1)

Proof. Suppose (1) holds. Then in each state, if the protocol should end at some leaf l, agent i can only get to a leaf in  $\{l' \in L : U_{-i}(l) \cap U_{-i}(l') \neq \emptyset\}$  by deviating, as it is the set of leaves that are attainable for him given that the types of the other agents are in  $U_{-i}(l)$ . Hence he will never be able to increase his payoff by deviating from the prescribed strategy, and the BDM is EPIC. Now suppose (1) is violated for some agent i, leaves l and l', and type  $u_i \in U_i(l)$ . Then in all states  $\{u_i\} \times (U_{-i}(l) \cap U_{-i}(l')) \neq \emptyset$ , agent i would be strictly better off following the strategy  $\sigma_i(u_i')$  for any type  $u_i' \in U_i(l')$ , which would violate EPIC. Note that we used the crucial assumption that all leaves l have non-empty legal domains U(l).

It immediately follows from Lemma 2 that information sets are irrelevant when considering ex post implementation.

**Corollary 1.** For every two BDMs  $\mathcal{B} = \langle \mathcal{P}, H, t \rangle$  and  $\mathcal{B}' = \langle \mathcal{P}, H', t \rangle$  that differ only in their information sets,  $\mathcal{B}$  is EPIC if and only if  $\mathcal{B}'$  is also EPIC.

Hence, for simplicity, we can restrict attention to PBDMs when we are in the EPIC setting.

Note that, by the Revelation Principle, any decision rule f that is implementable in an EPIC BDM with transfer rule  $t: U \to \mathbb{R}$  must be Dominant strategy Incentive Compatible (DIC) with this transfer rule, i.e., satisfy the following inequalities:

$$\forall u_i, u_i' \in U_i, \forall u_{-i} \in U_{-i} : u_i(f(u_i, u_{-i})) + t_i(u_i, u_{-i}) \ge u_i(f(u_i', u_{-i})) + t_i(u_i', u_{-i}). \tag{2}$$

Hence, all the EPIC-implementable rules we will consider in our private valuations setting are necessarily DIC-implementable. Also, using the above inequalities for  $u_i, u'_i \in U_i$  such that  $f(u_i, u_{-i}) = f(u'_i, u_{-i})$ , we see that  $t_i(u_i, u_{-i}) = t_i(u'_i, u_{-i})$ , and therefore the transfer to each agent i can be written in the form

$$t_i(u) = \tau_i(f(u), u_{-i}) \text{ for some } tariff \ \tau_i : X \times U_{-i} \to \mathbb{R} \cup \{-\infty\}.$$
 (3)

## 4.4 Bayesian Incentive-Compatibility

The concept of *Bayesian incentive-compatibility* means that for each input, the agents' strategies constitute a (interim) Bayesian Nash equilibrium given the probabilistic belief over the state.

**Definition 5.** Given probability distribution p over state space U,  $BDM \langle \mathcal{P}, H, t \rangle$  is Bayesian Incentive Compatible for p (or BIC(p) for short) if the strategies  $\sigma_1, \ldots, \sigma_N$  are measurable, and

$$\forall i \in I, \forall u_i \in U, \forall s_i' \in S_i: E_{u_{-i}}[u_i(x(g(\sigma(u)))) + t_i(g(\sigma(u)))|u_i] \ge$$

$$E_{u_{-i}}[u_i(x(g(s_i',\sigma_{-i}(u_{-i})))) + t_i(g(s_i',\sigma_{-i}(u_{-i})))|u_i].$$

In words, for every agent i and every type  $u_i \in U_i$ ,  $\sigma_i(u_i)$  maximizes the expected utility of the agent given the updated probability distribution  $p_{-i}(.|u_i)$  over the other agents' types  $u_{-i} \in U_{-i}$ , as long as they follow their prescribed strategies. We will assume that the types of the agents are independently distributed, i.e., that the probability distribution p over states is a product of the individual probability distributions over types  $p_i$ . Relaxation of this *independent types* setting will be discussed in Section 8.5.

By definition, Bayesian implementation is weaker than ex post implementation: if BDM  $\mathcal{B}$  is EPIC, then it is BIC(p) for every state distribution p. When the BDM is BIC(p), we say that it implements  $Fun(\mathcal{P})$  in BIC(p).

## 4.5 Incentivizability of Protocols

In standard mechanism design, according to the revelation principle, a decision rule is implementable for some equilibrium concept if and only if the direct revelation protocol for this decision rule can be incentivized with some transfers. Now we want to define incentivizability for general protocols.

**Definition 6.** A protocol  $\mathcal{P}$  with I agents and set of leaves L is EPIC-incentivizable if there is a transfer function  $t: L \to \mathbb{R}^I$  and a partition H of  $N \setminus L$  into information sets such that  $BDM \setminus \mathcal{P}, H, t \setminus S$  is EPIC.

A protocol  $\mathcal{P}$  with I agents and set of leaves L is BIC(p)-incentivizable if there is a transfer function  $t: L \to \mathbb{R}^I$  and a partition H of  $N \setminus L$  into information sets such that  $BDM \langle \mathcal{P}, H, t \rangle$  is BIC(p).

By Corollary 1, we can restrict attention to PBDMs and we can determine if the protocol is EPIC-incentivizable by determining if there is a solution to the system of linear inequalities described in the Lemma 2. Also, given a protocol, transfers satisfying EPIC can be effectively computed using the typical methods for linear programming (e.g., simplex method).

By definition, a protocol is EPIC-incentivizable only if it computes an EPIC-implementable decision rule. However, the converse is not true: a protocol computing an EPIC-implementable decision rule need not be EPIC-incentivizable.

Example 1. There are two agents and one indivisible object, which can be allocated to either agent. The two agents' valuations (utilities from receiving the object) lie in type spaces  $U_1 = \{1, 2, 3, 4\}$  and  $U_2 = [0, 5]$  respectively (their utilities for not receiving the object are normalized to zero). An efficient (and hence EPIC-implementable) allocation f of the object (i.e., to the agent with the higher valuation) can be computed with the following protocol  $\mathcal{P}_0$ : Agent 1 sends his type  $u_1$  (using  $\log_2 4 = 2$  bits), and then Agent 2 outputs an efficient allocation  $f(u) \in \{1,2\}$  (using 1 bit). Suppose in negation that the protocol computes transfer rule  $t_1$  that satisfies the ex post incentives of Agent 1. Given the information revealed during the execution of the protocol,  $t_1(u)$  can depend only on  $u_1$  and f(u). So it must take the form  $t_1(u) = t'_1(u_1, f(u))$  for some  $t'_1 : U_1 \times X \to \mathbb{R}$ . However, by (3), EPIC requires that  $t_1$  satisfy  $t_1(u) = \tau_1(f(u), u_2)$  for some  $\tau_1 : X \times U_2 \to \mathbb{R}$ . Hence  $t_1(u_1, u_2) = t_1^*(f(u))$  for some  $t_1^* : X \to \mathbb{R}$ . But then if  $t_1^*(1) - t_1^*(2) \le 2.5$ , Agent 1 would want to deviate in state  $(u_1, u_2) = (3, 3.5)$  to announcing  $u'_1 = 4$  and getting the object, while if  $t_1^*(1) - t_1^*(2) > 2.5$ , Agent 1 would want to deviate in state  $(u_1, u_2) = (2, 1.5)$  to announcing  $u'_1 = 1$  and not getting the object. In fact, it can be shown in this example that no 3-bit protocol computing an efficient decision rule is incentivable, hence the communication cost of selfishness is positive.

By definition, a protocol is BIC(p)-incentivizable only if it computes a BIC(p)-implementable decision rules. However, the converse is not true: a protocol computing a BIC(p)-implementable decision rule need not be BIC(p)-incentivizable.

**Example 2.** There are two agents. Agent 1 is always indifferent over outcomes and his type is a number  $m \in \{1, 2, 3, 4\}$ . Agent 2's type is a bit  $b \in \{0, 1\}$  and a string w of length 4 having the same number 4/2 = 2 of 0's and 1's. The decision rule dictates outcome  $x = w(m) \in X = \{0, 1\}$ 

<sup>&</sup>lt;sup>8</sup>To see this, we can check by exhaustion that there are only two 3-bit protocols computing an efficient decision rule: the one described in the example, and the one in which (i) Agent 1 first sends one bit about his valuation, (ii) Agent 2 says whether he already knows an efficient decision, (iii) If Agent 2 said yes, he announces the decision, and otherwise Agent 1 announces an efficient decision. In this protocol, the EPIC constraints of Agent 2 cannot be satisfied with any transfers, by a similar argument to that in Example 1.

and the utility function of Agent 2 gives him payoff  $(x + b) \mod 2$  (i.e., the "xor" of x and b) for each outcome x. The types are distributed uniformly over their domains. The decision rule satisfies BIC(p) without transfers, since whatever is Agent 2's announcement, he receives the same expected utility of 1/2. However, consider the following protocol  $\mathcal{P}_1$  that computes this BIC(p)-implementable decision rule: Agent 1 outputs m and then, Agent 2 outputs x = w(m). Note that any BDM constructed with this protocol will need to give Agent 2 perfect information, as he needs to be able to distinguish all of the announcements of Agent 1. However, Agent 2 will report the value of x that maximizes his total payoff  $(x + b) \mod 2 + t_2(m, x)$ , where  $t_2$  is his transfer function in the protocol (the actual leaf depends only on m and x). For at least one of the values of b, his total payoff will have a unique maximum, and the value of x that achieves it does not depend on w. Hence  $\mathcal{P}_1$  is not BIC(p)-incentivizable.

On the other hand, there are examples in which the communication cost of selfishness is zero. For example, for a decision rule f that is EPIC implementable with some transfer rule t that is a function of only the outcome, <sup>9</sup> any protocol computing the f is EPIC-incentivizable. Also, as we will see in Section 7.1, all protocols that compute an EPIC-implementable decision rule are BIC(p)-incentivizable for every state distribution p.

## 4.6 Incentive Communication Complexity

We cannot necessarily use, in an EPIC BDM, the shallowest communication protocol that computes the EPIC-implementable decision rule since it may not be EPIC-incentivizable. Likewise, we cannot necessarily use, in a BIC(p) BDM, the shallowest communication protocol that computes the BIC(p)-implementable decision rule since it may not be BIC(p)-incentivizable. So the need to satisfy the incentives may create an additional communication cost in both implementations.

**Definition 7.**  $CC^{EPIC}(f)$ , the ex post incentive communication complexity of a decision rule f, is the depth  $d(\mathcal{B})$  of the shallowest BDM  $\mathcal{B}$  that implements f in EPIC.

 $\mathrm{CC}_p^{BIC}(f)$ , the Bayesian incentive communication complexity of a decision rule f with state distribution p, is the depth  $d(\mathcal{B})$  of the shallowest BDM  $\mathcal{B}$  that implements f in BIC(p).

Note that since Bayesian implementation is weaker than ex post implementation, for every EPIC-implementable rule and state distribution p,  $CC_p^{BIC}(f) \leq CC^{EPIC}(f)$ .

Formally, in the expost setting, the communication cost of selfishness (overhead for short) is the difference between CC(f) and  $CC^{EPIC}(f)$ . Likewise, in the Bayesian setting with state distribution p, the overhead is the difference between CC(f) and  $CC^{BIC}_p(f)$ .

# 5 Overhead for Ex Post Implementation

## 5.1 Characterization of the Overhead

For ex post implementation, according to Corollary 1, we never need to hide information from agents so that, as we noticed before, the restriction to considering only PBDMs is without loss of generality. Intuitively, this is because ex post implementation requires that every agent follows his prescribed strategy even if he knows the types of the other agents (i.e., the complete state). Hence the overhead

<sup>&</sup>lt;sup>9</sup> For example, suppose we have one object and two agents, the object goes to Agent 1 if his valuation is more than a, otherwise, it goes to Agent 2 if his valuation is more than b, and is not sold if neither condition is satisfied. This decision rule is EPIC implementable using transfers that depend only on the allocation of the object.

comes only from the need of revealing additional information to compute the right transfers for each agent. Therefore, the incentive communication complexity of ex post implementation can be seen as the communication cost of the cheapest protocol among all protocols that compute an EPIC social choice rule (i.e., decision rule + transfer rule) implementing our decision rule.

Formally, let T be the set of all possible transfer rules  $t: U \to \mathbb{R}^I$  that satisfy the expost incentive constraints for a given decision rule f, and let the function  $f_t: U \to X \times \mathbb{R}^I$  be defined as  $f_t(u) = (f(u), t(u))$ . Then,  $\{f_t: t \in T\}$  is exactly the set of EPIC social choice rules implementing f, and  $CC^{EPIC}(f)$  is exactly  $\min_{t \in T} CC(f_t)$ . As such, it is radically different from the communication complexity of any given function or relation, as the set of acceptable transfers at one leaf depends on the transfers given at other leaves.

## 5.2 Exponential Upper Bound on the Overhead

We find that the communication cost of selfishness in this case can be bounded by a function of the original communication complexity of the decision rule.

By the revelation principle, given an EPIC-implementable decision rule, if there is full revelation of the types, then we can compute a transfer rule that makes the incentive constraints hold. But is full revelation necessary to find acceptable transfers? For example, if the state space is infinite, but the decision rule has finite communication complexity, could it be the case that we need infinite communication to find acceptable transfers? The answer is no, it is not possible. In fact, given a protocol that computes an implementable decision rule, it is sufficient to reveal what each agent would do at any node. Note that it is equivalent to full revelation of the strategies in the protocol, and that with SBDMs, the strategies are always fully revealed. Hence:

**Proposition 1.** Given an EPIC-implementable decision rule f, every simultaneous communication protocol that computes f is EPIC-incentivizable.

Proof. Consider any simultaneous communication protocol  $\mathcal{P}$  with information sets H such that no agent learns anything during any execution of  $\mathcal{P}$ , and pick any transfer rule  $t: U \to \mathbb{R}^I$  that satisfies the incentive constraints. Note that there is a one-to-one correspondence  $\delta: L \to \prod_{i \in I} S_i$ , where  $\prod_{i \in I} S_i$  is the set of all possible profiles of strategies. Then fix any function  $\overline{\sigma}: S \to U$  such that  $\sigma \circ \overline{\sigma}$  is the identity function on S. This is possible because  $\sigma$  is surjective, i.e.,  $\sigma(U) = S$ . Now the SBDM  $\langle \mathcal{P}, H, t' \rangle$  is EPIC with  $t': L \to \mathbb{R}^I$  defined as  $t' = t \circ \overline{\sigma} \circ \delta$ , because, from the inequalities (2), it follows that for every agent i, different types in  $U_i$  that are prescribed the same strategy in  $\mathcal{P}$  can be given the same transfer for every  $u_{-i} \in U_{-i}$ .

Note that the resulting SBDM also satisfies the stronger Dominant strategy Incentive Constraints (DIC) since every strategy in the strategy space is used by some type.

Our restriction to private valuations is crucial here. There are EPIC decision rules with interdependent valuations that do not satisfy the above property. We discuss the interdependent valuations case in more details in Section 8.4.

Corollary 2. If f is an EPIC-implementable decision rule, then:  $CC^{EPIC}(f) \leq 2^{CC(f)} - 1$ 

*Proof.* For any protocol  $\mathcal{P}$  that achieves the lower bound CC(f), by Lemma 1, there is a simultaneous communication protocol  $\mathcal{P}'$  computing the same decision rule such that  $d(\mathcal{P}') \leq 2^{d(\mathcal{P})} - 1 = 2^{CC(f)} - 1$ . By Proposition 1,  $\mathcal{P}'$  is EPIC-incentivizable, which proves the result.

This upper bound shows that the communication cost of selfishness is not unbounded but is at most exponential. In particular, it shows that a decision rule f can be implemented in an EPIC BDM if and only if f is an EPIC-implementable decision rule that can be computed with finite communication. We do not need the stronger requirement that the state space U be finite.

This upper bound can be improved upon by eliminating those strategies in the simultaneous protocol that are not used by any type.

**Example 3.** Consider the setting in Example 1. Protocol  $\mathcal{P}_0$  has depth 3, so by Proposition 2, there is a protocol of depth  $2^3 - 1 = 7$  that is EPIC-incentivizable. But we can go further: Agent 1 needs only 4 strategies in  $\mathcal{P}_0$  (one for each of his types) out of the  $2^3 = 8$  possible strategies, and Agent 2 needs only 5 strategies out of  $2^4 = 16$ , each of the 5 strategies being described by a threshold of Agent 1's announcement below which Agent 2 takes the good. Hence, full description of such strategies takes  $\lceil \log_2 4 \rceil + \lceil \log_2 5 \rceil = 5$  bits. So there is a protocol of depth 5 that is EPIC-incentivizable.

It is unknown, however, whether there is an EPIC-implementable decision rule f such that  $CC^{EPIC}(f)$  is even close to this exponential bound of  $2^{CC(f)}$ . In particular, an open problem is to determine the highest attainable upper bound, and to determine if there are canonically hard instances, in the spirit of [Feigenbaum *et al.*, 2002].

**Open Problem 1.** Are there canonically hard EPIC-implementable decision rules for which the incentive communication complexity of decision rule f,  $CC^{EPIC}(f)$ , is much higher than the communication complexity of f, CC(f)? How high can the communication cost of selfishness be for expost implementation?

In the meantime, we consider some low overhead cases in Sections 7.2, 7.3 and 7.4. Also, we show in Section 8.1 that the communication cost of selfishness is arbitrarily high when we consider the average-case communication cost measure.

# 6 Overhead for Bayesian Implementation

## 6.1 Characterization of the Overhead

Intuitively, with Bayesian implementation, agents might be better off lying if they have too many contingent deviations, so the overhead comes from the need to hide information from the agents, and the restriction to PBDM is not without loss of generality. Given a protocol, we can minimize the set of deviations by having maximally large information sets, which minimizes the number of information sets for each agent while satisfying perfect recall and still giving each agent enough information to compute his prescribed move at each step. E.g., SBDMs have maximally large information sets, as each agent never learns anything during the execution of the protocol. Hence a BDM will be BIC(p) only if it stays BIC(p) when we replace its information sets by maximally large ones. The idea of hiding as much information as possible to maximize incentives can be traced back to Myerson's "communication equilibrium" [Myerson, 1991] where a mediator receives the types of the agents and tells each of them only what his prescribed move is. Likewise, in a protocol, if at each node  $n \in N_i$  agent i learns only what his prescribed move function is (the function  $\sigma_i(.)(n)$  from types  $U_i$  to moves in  $\{0,1\}$ ), it yields maximally large information sets and maximizes the agent's incentives.

On the other hand, computing transfers is not a problem with Bayesian implementation, as the following result suggests:

**Lemma 3.** Let  $\mathcal{P}$  be a BIC(p)-incentivizable protocol computing decision rule f, and let  $\mathcal{P}'$  be the protocol we get from  $\mathcal{P}$  when, in all states, we stop the execution of  $\mathcal{P}$  once the outcome is known. Then  $\mathcal{P}'$  is also BIC(p)-incentivizable.

*Proof.* It follows by induction using the following statement: Let  $\mathcal{P}$  be a BIC(p)-incentivizable protocol computing decision rule f, and let l and l' be two sibling leaves that are the left and right children of some node n, and that satisfy x(l) = x(l') = x, then the protocol  $\mathcal{P}'$  we get from  $\mathcal{P}$  by removing l, l' and adding n as a leaf with outcome x(n) = x is also BIC(p)-incentivizable. Consider a BIC(p) BDM  $\mathcal{B} = \langle \mathcal{P}, H, t \rangle$  and let  $h \in H_i$  be the information set that contains n. We construct from it the BIC(p) BDM  $\mathcal{B}' = \langle \mathcal{P}', H', t' \rangle$  in the following way. First note that the incentives of all agents except agent i can be satisfied in  $\mathcal{B}'$  by giving them at leaf n the expected value of their transfers at l and l', so we will focus on satisfying the incentives of agent i. If  $h = \{n\}$ , note that we must have  $t_i(l) = t_i(l')$  (because both leaves have a non-empty domain) and we define  $H' = H \setminus h$ ,  $t_i'(n) = t_i(l)$ , with the same values as t everywhere else. Clearly,  $\mathcal{B}'$  satisfies the incentives of agent i since it is strictly the same for him to be at n, l or l' in  $\mathcal{B}$ . Now, if h contains more nodes, let  $q \in ]0,1[$  be the probability for agent i that he is at node n given that he is at information set h. First note that BIC(p) is satisfied for the BDM  $\langle \mathcal{P}, H, t'' \rangle$ , where t'' is the same as t except that  $t_i''(l) = t_i''(l') = t_i(l)$  and for every leave l'' that can be reached after a right move from a node in  $h \setminus \{n\}$ ,  $t_i''(l'') = t_i(l'') + (t_i(l') - t_i(l))q/(1-q)$ . Hence, by defining  $t_i'(n) = t_i(l)$  and  $t_i'$  as  $t_i''$ everywhere else, the incentives constraints of agent i are satisfied in  $\mathcal{B}'$ .

Loosely speaking, if the BDM is BIC(p) and if at some node n we have enough information to compute the outcome, we will not need to get more information from the agents to satisfy their incentives, and there is a transfer t(n) for which we can "stop" the protocol while keeping the agents' incentive constraints satisfied.

Note that the independence of types is crucial for the above property (although it extends to interdependent valuations): there are BIC(p) decision rules with correlated types that do not satisfy it. We will consider correlated types in more details in Section 8.5.

In general, it is hard to check if a protocol is BIC(p)-incentivizable. Also, given a BDM, it is hard to check if it is BIC(p) since we need to consider the expected payoff that each agent would get for each possible deviation. It seems that the only way to check this is by dynamic programming: starting from the leaves, we compute what is the "expected value" of every node of the protocol for every type of each agent. Then we can check that at each information set, the agent maximizes his expected payoff by following the prescribed move for every legal type.

However, a useful sufficient condition for a BDM to be BIC(p) is that no information that an agent i can receive during the execution of the BDM (whatever strategy  $s \in S_i$  he uses) can ever make his prescribed strategy suboptimal. Formally:

**Proposition 2.** A protocol  $\mathcal{P}$  computing decision rule f is BIC(p)-incentivizable with information sets H if there is a transfer rule  $t: U \to \mathbb{R}^I$  such that, for every  $i \in I, h \in H_i$ , the decision rule f with transfer  $t_i$  satisfies  $BIC(p^h)$  for agent i with the updated distribution  $p^h$  over U at h, i.e.,

$$\forall u_i, u_i' \in U_i, E_{u_{-i}}[u_i(f(u)) + t_i(u)|h] \ge E_{u_{-i}}[u_i(f(u_i', u_{-i})) + t_i(u_i', u_{-i})|h].$$

Proof. It follows from the fact that the BDM  $\langle \mathcal{P}, H, t' \rangle$  is BIC(p) with  $t'_i(l) = E_u[t_i(u)|u \in U(l)]$ , where u follows state distribution p. For any possible deviation, and whatever information he learns about the state, agent i will always (weakly) regret not having been truthfull. Hence he cannot have a profitable deviation.

By Lemma 3, the transfer rule used in the proposition need not be computed by the protocol. We only need its existence, and the protocol will compute some transfer rule (e.g., t' defined as  $t'_i(l) = E_u[t_i(u)|u \in U(l)]$ ) that will satisfy the incentives of the agents at each step. Also, note that Proposition 2 uses the interim definition of Bayesian incentive-compatibility, as each agent i must still be maximizing his expected utility at information set h when his type  $u_i$  has probability  $p_i^h(u_i) = 0$ . Finally if we want to check that a protocol is BIC(p)-incentivizable with some information sets using Proposition 2, we need to consider, for every  $i \in I$ , that f is  $BIC(p^h)$  with the same fixed transfer rule t only for the information sets of agent i that can reach a leaf without going through any other information set of agent i. This will immediately imply that the property holds for the information sets of agent i that are above.

## 6.2 Exponential Upper Bound on the Worst-Case Overhead

Much in the same spirit as with ex post implementation, the following result holds:

Corollary 3. Given a BIC(p)-implementable decision rule f, any simultaneous communication protocol that computes f is BIC(p)-incentivizable.

*Proof.* Consider the information sets H that an SBDM with this protocol would necessarily have. With these information sets, since no agent learns anything about the state during the execution of the protocol, the result follows immediately from Proposition 2 and from the definition of BIC(p)-implementability.

Corollary 4. If f is a BIC(p)-implementable decision rule, then:  $CC_p^{BIC}(f) \leq 2^{CC(f)} - 1$ .

*Proof.* Given the protocol  $\mathcal{P}$  that achieves the lower bound CC(f), by Lemma 1, there is a simultaneous communication protocol  $\mathcal{P}'$  that computes the same decision rule such that  $d(\mathcal{P}') \leq 2^{d(\mathcal{P})} - 1 = 2^{CC(f)} - 1$ . By Corollary 3,  $\mathcal{P}'$  is incentivizable, which proves the result.

This upper bound shows that the communication cost of selfishness is not unbounded but is at most exponential. In particular, it shows that a decision rule f can be implemented in a BIC(p) BDM if and only if f is a BIC(p)-implementable decision rule that can be computed with finite communication. We do not need the stronger requirement that the state space U be finite.

#### 6.3 The Upper Bound is Tight for Bayesian Implementation

Unlike the ex post implementation, where we do not know if the upper bound is tight, we have here an example that achieves the upper bound for the Bayesian case.

**Proposition 3.** For any large M, there is a BIC(p)-implementable decision rule  $f: U \to X$  such that  $CC(f) \leq 2\lceil \log_2 M \rceil$  but  $CC_p^{BIC}(f) \geq 0.5 \log_2 \binom{M}{M/4}$ , with uniform state distribution p.

Proof. See Appendix A.1. 
$$\Box$$

By Stirling's formula,  $0.5 \log_2 \binom{M}{M/4}$  is asymptotically of order M, which is exponentially higher than  $\lceil \log_2 M \rceil$ . This shows that our exponential upper bound is essentially tight. The above proposition is proved by formalizing the following example:

**Example 4.** There are M possible decisions (outcomes) and M possible consequences. A Manager has a uniformly distributed consequence  $c \in \{1, ..., M\}$  he would like to achieve. An Expert knows a one-to-one correspondence between decisions and consequences, but also has a private utility

function over consequences. The goal is to compute the decision that yields the Manager's desired consequence according to the correspondence known by the Expert. With an honest Expert, the communication complexity CC(f) is at most  $2\lceil \log_2 M \rceil$  bits: the Manager reports the consequence he wants, and the Expert reports the corresponding decision. With a selfish Expert, a direct revelation mechanism satisfies the Expert's Bayesian incentives without using transfers, since the Expert faces the same uniform distribution over consequences whatever he reports. The communication cost of the Expert's reporting his correspondence is of order  $\log_2 M! \sim M \log_2 M$  (the Expert need not be asked to reveal his preferences). Furthermore, we prove that any mechanism satisfying the Expert's Bayesian incentive constraints has a communication cost that cannot be significantly lower – it must be at least of order M, which is still exponential in CC(f). Intuitively, any simpler mechanism will be susceptible of Expert influence, as he will infer how his knowledge will be used and bias his report according to his preferences. This is true even if the mechanism can use transfers. (Note that we assume that the realized consequence is not contractible.—e.g., it is not observed by the Manager until after the mechanism is finished.)

# 7 Low-Overhead Cases for EPIC-Implementable Decision Rules

While we should not expect the communication cost of selfishness to be low in general, we identify cases where it is reasonable.

In the first section, we show that there is no overhead in the Bayesian setting with an EPIC-implementable decision rule.

In the ex post setting, we consider the case where the decision rule f is efficient, i.e., the decision rule always chooses an outcome x that maximizes the sum of utilities  $\sum_{i \in I} u_i(x)$ . The applicability of these results goes beyond efficient decision rules: under some natural conditions like irrelevance of independent alternatives, or with unrestricted valuation space (see [Lavi et al., 2003]), any EPIC-implementable decision rule maximizes a non-negative affine combination of the agents' utilities, which can be interpreted as efficiency upon rescaling the utilities and adding a fictitious agent with known utility. We consider efficiency in general and also when agents are single-parameter, i.e., they have a non-zero valuation on a fixed set of equally desirable outcomes.

In the last two sections, we finally consider the case where there are only two agents and either a fixed utility precision or a fixed number of outcomes, for all EPIC-implementable decision rules.

### 7.1 No Overhead with Bayesian Implementation

In this case, which includes all efficient decision rules, there is no communication cost of selfishness.

**Proposition 4.** If  $\mathcal{P}$  is a protocol that computes an EPIC-implementable decision rule f, then  $\mathcal{P}$  is BIC(p)-incentivizable for every state distribution p.

*Proof.* This follows from Proposition 2 as the decision rule f is EPIC-implementable with some fixed transfer t, and hence f is also BIC(p)-implementable for every p with the same transfer t.  $\square$ 

This result is related to [Athey and Segal, 2005], which considers Bayesian implementation in a dynamic environment. The paper also shows that the above results can be extended to the case where types are correlated and that the property of budget balance (i.e.,  $\forall u, \sum_{i \in I} t_i(u) = 0$ ), which is often the motivation behind the use of Bayesian implementation, can be satisfied in BDMs as long as the types are independently distributed.

Proposition 4 also shows that there is no communication cost of selfishness with respect to the average-case communication cost measure (defined in Section 8.1) when we use Bayesian implementation with an EPIC-implementable decision rule.

## 7.2 Efficient Decision Rule: Finite Utility Range

It is well known that any efficient decision rule is EPIC-implementable, and can even be implemented in dominant strategies just by giving to each agent the sum of other agents' utilities from the computed outcome (as in the VCG mechanism). Following the same idea, starting with any protocol computing an efficient decision rule f, we can satisfy EPIC by having the agents report their utilities from the outcome computed by the protocol, and then transfer to each agent the amount equal to the sum of the reported utilities of the other agents. This approach dates back to Reichelstein [Reichelstein, 1984] and was more recently used in [Athey and Segal, 2005]. This approach works well when agents' utilities are given in discrete multiples:

**Definition 8.** The utility functions in  $U = (u_1, ..., u_I)$  have discrete range with precision  $\gamma$  if  $u_i(x) \in \{k2^{-\gamma} : k = 0, ..., 2^{\gamma} - 1\}.$ 

In this case, we can modify any efficient protocol to make it EPIC-incentivizable as follows: At each leaf l, let each agent reports his utility using  $\gamma$  bits, which will determine the transfers of the other agents. Thus, we have

**Proposition 5.** For a utility function space U with discrete precision- $\gamma$  range, and an efficient decision rule f,

$$CC^{EPIC}(f) \le CC(f) + I\gamma$$

Thus, the communication cost of selfishness is at most  $\gamma I$  bits.

Furthermore, even if utility range is continuous but bounded (without loss of generality lies in [0, 1], up to a rescaling of utilities), we can use the same approach to get  $\epsilon$ -approximate EPIC (i.e., an agent will follow the prescribed strategy unless he can get more than  $\epsilon$  by deviating). Indeed, it is sufficient for each agent to output his utility rounded down to a multiple of  $\epsilon/(I-1)$  to construct transfers that are sufficiently near VCG transfers to ensure obedience. This will take at most  $\lceil \log_2 ((I-1)/\epsilon) \rceil$  bits by each agent, which bounds above the communication cost of selfishness.

In some cases we cannot achieve full efficiency, e.g., for lack of sufficient computational resources to compute the outcome, and we must settle for a decision rule f that is  $\epsilon'$ -efficient, i.e., satisfies

$$\sum_{i \in I} u_i(f(u)) \ge \max_{x \in X} \sum_{i \in I} u_i(x) - \varepsilon' \ \forall u \in U.$$

The above argument for  $\epsilon$ -EPIC implementation is adapted to such decision rules as long as  $0 \le \epsilon' < \epsilon$ : It is sufficient for each agent to output his utility rounded down to a multiple of  $(\epsilon - \epsilon')/(I - 1)$ , which takes at most  $\lceil \log_2 ((I - 1)/(\epsilon - \epsilon')) \rceil$  bits. To summarize:

**Proposition 6.** If U is a utility function space with utility range contained in [0,1], and decision rule f is  $\varepsilon'$ -efficient, and the solution concept is  $\varepsilon$ -EPIC with  $0 \le \epsilon' < \epsilon$  then

$$CC^{EPIC}(f) \le CC(f) + I\lceil \log_2 (I-1)/(\epsilon - \epsilon') \rceil$$

So in general, for efficient and approximately efficient decision rules, the communication cost of selfishness is bounded above by a number that does not depend on the communication complexity of the decision rule, and that is low if the utility range is discrete or if approximate EPIC is allowed.

## 7.3 Efficient Decision Rule: Single-Parameter Agents

**Definition 9.** Agent i is a single-parameter agent if his type is defined by some real  $v_i \in V_i \subset \mathbb{R}$  such that his utility function takes the form  $u_i(x) = v_i a_i(x) + b_i(x)$ , for some functions  $a_i, b_i : X \to \mathbb{R}$ .

With single-parameter agents, for efficient rules, it turns out that we can have all the relevant information about each valuation to find acceptable transfers with an overhead that is at most linear in the communication complexity of the decision rule (and this is regardless of the utility range).

**Proposition 7.** Consider an environment with I single-parameter agents. Let f be an  $\varepsilon$ -efficient decision rule. Then:

$$CC^{EPIC}(f') \leq I \cdot (CC(f) + 1)$$

for some  $\varepsilon$ -efficient decision rule f'.

*Proof.* Consider a protocol  $\mathcal{P}$  computing f. For each agent i, let  $\underline{u}_i(l) = \inf U_i(l)$  and  $\overline{u}_i(l) = \sup U_i(l)$ . Given the linear utilities of SP agents, the outcome x(l) must be  $\varepsilon$ -efficient on  $\prod_{i \in I} [\underline{u}_i(l), \overline{u}_i(l)]$ .

The threshold points  $\underline{u}_i(l)$ ,  $\overline{u}_i(l)$ ,  $l \in L$  partition agent *i*'s type space  $U_i$  into at most 2|L| intervals. Consider now a simultaneous communication protocol  $\mathcal{P}'$  in which each agent *i* reports which of these intervals his type lies in, and the outcome x(l) for a leaf  $l \in L$  for which  $u \in \prod_{i \in I} [\underline{u}_i(l), \overline{u}_i(l)]$  is

implemented. The new protocol computes an  $\varepsilon$ -efficient outcome with each agent sending at most  $\log_2(2|L|) \leq 1 + d(\mathcal{P})$  bits. Furthermore, since it is a simultaneous communication protocol,  $\mathcal{P}'$  is EPIC-incentivizable by Proposition 1.

For example, in the problem of allocating one indivisible object among I agents without externalities, the agents are single-parameter agents, and the theorem implies that selfishness multiplies the communication complexity of achieving a given efficiency approximation by at most  $I^{10}$ .

## 7.4 Two Agents with Fixed Utility Precision

Recall from (3) that when decision rule f is EPIC-implementable with transfer rule t, then the transfer  $t_i$  to agent i can be written as  $t_i(u) = \tau_i(f(u), u_{-i})$ . Furthermore, if the utilities have discrete range, then we could restrict attention to discrete transfers. With two agents, agent -i can output the transfer at the end of any protocol computing f(u), and so we obtain a BDM implementing f. This argument yields

**Proposition 8.** Suppose that I = 2 and that the utility function space U has discrete range with precision  $\gamma$ . Then for every EPIC-implementable decision rule f,

$$CC^{EPIC}(f) \le CC(f) + 2(\gamma + 1)$$

Proof. We can fix  $\tau_i(x_0, u_{-i}) = 0$  for every  $u_{-i}$  for an abritrary fixed outcome  $x_0 \in X$ . Then EPIC implies that  $|\tau_i(x, u_{-i})| \leq 1 - 2^{-\gamma}$ . Furthemore, since utilities have discrete range with precision  $\gamma$ , we can round down all transfers to multiples of  $2^{-\gamma}$  while preserving EPIC. Then reporting such a transfer takes  $1 + \gamma$  bits.

 $<sup>^{10}</sup>$ This result could also be derived using a theorem in [L. Blumrosen and Segal, ] that shows that any sequential communication in this setting can be replaced with simultaneous communication with multiplying the complexity by at most I, using the fact that any simultaneous communication protocol computing an EPIC-implementable decision rule is EPIC-incentivizable.

Note that we could get a parallel result for a continuous but bounded utility range for  $\epsilon$ -EPIC, since we can round down the transfers to multiples of  $\epsilon^{-1}$ . Furthermore, the argument can be adapted to  $\epsilon'$ -EPIC-implementable decision rules, provided that  $0 \le \epsilon' < \epsilon$ , and so the transfer can be rounded down to a multiple of  $(\epsilon - \epsilon')$ .

**Proposition 9.** Let I=2 with utility function space U with utility range contained in [0,1]. If decision rule f is  $\varepsilon'$ -EPIC-implementable, and the solution concept is  $\varepsilon$ -EPIC with  $0 \le \epsilon' < \epsilon$ , then

$$CC^{EPIC}(f) \le CC(f) + 2[-\log_2(\epsilon - \epsilon') + 1]$$

# 7.5 Two Agents with Fixed Number of Outcomes

**Definition 10.** For any EPIC-implementable decision rule  $f: U \to X$ , the revenue maximizing tariff  $\tau$  with reservation outcome  $x_0 \in X$  is the tariff such that, for every agent  $i \in I$  and type  $u_{-i} \in U_{-i}$ ,  $\tau_i: X \times U_{-i} \to \mathbb{R} \cup \{-\infty\}$  satisfies  $\tau_i(x_0, u_{-i}) = 0$  while for every outcomes  $x \in X$ ,  $\tau_i(x, u_{-i})$  is as small as possible under the EPIC inequalities.

The above characteristics uniquely define the tariff (we can always minimize the tariffs of all  $x \in X \setminus \{x_0\}$  at once) and the corresponding transfer rule  $t_i(u) = \tau_i(f(u), u_{-i})$ .

For every agent i and  $u_{-i} \in U_{-i}$ , we can construct a tree  $T(u_{-i})$  in the following way: The root is labeled with outcome  $x_0$  and for each outcome  $x \in X$  that binds  $\tau_i(x_0, u_{-i})$  for some type  $u_i \in U_i$  (i.e.  $f(u_i, u_{-i}) = x$  and  $u_i(x_0) + \tau_i(x_0, u_{-i}) = u_i(x) + \tau_i(x, u_{-i})$ ), we construct a child labeled with both x and one of the binding types  $u_i$ . We then construct the children of these new nodes labeled with outcome x recursively by finding every outcome x' that binds the tariff for x. We avoid loops in the process by discarding outcomes that already have a node in the tree. We continue the construction until all the possible outcomes with  $u_{-i}$ , i.e., all  $x \in \{f(u_i, u_{-i}) : u_i \in U_i\}$ , have a corresponding node in the tree.

With any outcome  $x \in X$ , agent -i having type  $u_{-i}$  can output the label of the parent of the node that contains  $x \in X$ , i.e., the outcome x' and a type  $u_i \in U_i$  of agent i that satisfies f(u) = x' and binds the value of  $\tau_i(x, u_{-i})$ . With this information, the protocol can compute  $\tau_i(x, u_{-i}) - \tau_i(x', u_{-i}) = u_i(x') - u_i(x)$ . Given an initial protocol  $\mathcal{P}$  that computes the decision rule, this can be done by having agent -i output the unique leaf l in  $\mathcal{P}$  such that x(l) = x' and  $(u_i, u_{-i}) \in U(l)$ . In this case  $\tau_i(x, u_{-i}) - \tau_i(x', u_{-i}) = \min_{u_i \in U_i(l)} u_i(x') - u_i(x)$ , and this difference can be computed by the protocol. Finally, by repeating this procedure recursively with x', and and its ancestors, agent -i will eventually get to the root and have  $x_0$  binding (at this last step, if  $x_0$  cannot be implemented with  $u_{-i}$ , agent -i needs to output the leaf with the outcome x and the type  $u_i$  that bind the tariff of  $x_0$ ). At this point, the protocol can compute the value of the tariff for x by adding the differences computed along the way. Hence:

**Proposition 10.** Let I=2 with EPIC-implementable decision rule  $f:U\to X$ , then:

$$CC^{EPIC}(f) \le (2Card X - 1)CC(f)$$

Proof. Given a shallowest protocol  $\mathcal{P}$  computing f, we construct  $\mathcal{P}'$  by having the agents first execute  $\mathcal{P}$ , and then each of the two agents outputs the sequence of leaves that binds the revenue maximizing tariff of each outcome, until each one gets to the (arbitrarily chosen) reservation outcome  $x_0$  (or the leaf with outcome and type that binds  $x_0$ ). Each leaf needs at most CC(f) bits to be described, and there is at most one leaf that needs to be described for each agent and outcome (except for the outcome of the run). Hence the depth of  $\mathcal{P}'$  is at most: CC(f) + 2(Card X - 1)CC(f) = (2Card X - 1)CC(f). To conclude, note that  $\mathcal{P}'$  computes the revenue maximizing tariff, and hence it is EPIC-incentivizable.

## 8 Extensions

In this section, we consider several extensions of our analysis, namely to average-case communication cost, decision correspondence, dominant strategy implementation, interdependent valuations, and correlated types. We indicate when our previous results continue to hold, and when they need to be modified.

## 8.1 Average-Case Communication Cost

The communication cost measure used so far is the number of bits sent during the execution of a protocol in the *worst case*. We may also be interested in the *average-case* communication cost, given some probability distribution over the inputs:

**Definition 11.** If  $u \in U$ , let  $d(\mathcal{P}, u)$  be the depth of the (unique) leaf l in protocol  $\mathcal{P}$  such that  $u \in U(l)$ . For each state probability distribution p over U, we define the average communication cost of  $\mathcal{P}$  as  $ACC_p(\mathcal{P}) = E_u[d(\mathcal{P}, u)]$ . Furthermore, given a function  $f: U \to X$ , we define the average communication complexity of f given state distribution p as  $ACC_p(f) = \min_{\mathcal{P}: Fun(\mathcal{P}) = f} ACC_p(\mathcal{P})$ .

 $ACC_p^{EPIC}(f)$  and  $ACC_p^{BIC}(f)$ , the average ex post and Bayesian incentive communication complexity of a decision rule f with state distribution p, is the minimal average communication cost  $ACC_p(\mathcal{B})$  over every BDMs  $\mathcal{B}$  that implements f in EPIC or BIC(p) respectively.

When we consider the average-case communication costs, the communication cost of selfishness is the difference between  $ACC_p(f)$  and  $ACC_p^{EPIC}(f)$  with expost implementation, and it is the difference between  $ACC_p(f)$  and  $ACC_p^{BIC}(f)$  with Bayesian implementation.

Note that for every protocol  $\mathcal{P}$  and for every state distribution p,  $ACC_p(\mathcal{P}) \leq CC(\mathcal{P})$ . It follows immediately that for every decision rule f and every state distribution p:  $ACC_p(f) \leq CC(f)$ ,  $ACC_p^{EPIC}(f) \leq CC^{EPIC}(f)$  and  $ACC_p^{BIC}(f) \leq CC_p^{BIC}(f)$ .

With ex post implementation, the communication cost of selfishness is unbounded, even with two agents, an efficient decision rule and the uniform state distribution.

**Proposition 11.** For every  $\alpha > 0$ , there exists an efficient decision rule f with two agents such that, given the uniform state distribution p:  $ACC_p(f) < 4$  but  $ACC_p^{EPIC}(f) > \alpha$ .

Proof Sketch. Consider the problem of allocating an indivisible object to one of the two agents, as in Example 1 above, but with the agents' types drawn independently from a uniform distribution over  $U_1 = U_2 = \{k2^{-\gamma} : k = 0, ..., 2^{\gamma} - 1\}$ . Let f be the "efficient" decision rule. f can be computed using a bisection protocol with an average-case communication cost of at most 4 bits, whatever is the precision  $\gamma$ . However, any BDM that implements f in EPIC must compute the exact valuation of at least one of the agents, with a positive probability (say, at least 1/32). This will take communication that is of the order of  $\gamma$  bits. See Appendix A.2 for the complete proof.  $\square$ 

Our average-case analysis can be extended to infinite protocols. While we have not formally defined such protocols, we can imagine that there some such protocols whose average-case cost is finite. E.g., we can use such a protocol to find an efficient allocation for an object between two agents whose valuations are uniformly distributed on [0,1] using only 4 bits on average. However, no protocol having a finite average-case communication cost is EPIC-incentivizable in this case.

The communication cost of selfishness is also unbounded for Bayesian implementation, even with only two agents:

**Proposition 12.** For any  $\alpha > 0$  and  $\epsilon > 0$ , there exists a BIC(p)-implementable decision rule f with two agents such that:  $ACC_p(f) < 1 + \epsilon$  but  $ACC_p^{BIC}(f) > \alpha$ .

*Proof.* Consider the rule f used to prove Proposition 3. The rules satisfies  $ACC_p(f) \leq CC(f) \leq 2\lceil \log_2 M \rceil$ , but also satisfies  $ACC_p^{BIC}(f) \geq 0.5 \log_2 \binom{M}{M/4}$ , as it is shown in Appendix A.1. Let us construct the following rule f' from the rule f, by extending the type of Agent 2 to include a bit b that is equal to 1 with probability  $0.5\epsilon/\lceil \log_2 M \rceil$ , and by adding an outcome  $x_0$  that always gives 0 utility to every agent for every type. f' dictates  $x_0$  whenever b=0 and dictates the same outcome as f whenever b=1. We get by construction:

$$ACC_p(f') \le 1 \cdot (1 - 0.5\epsilon / \lceil \log_2 M \rceil) + 2\lceil \log_2 M \rceil \cdot 0.5\epsilon / \lceil \log_2 M \rceil < 1 + \epsilon$$

And we also get that  $ACC_p^{BIC}(f') \ge 1/4 \cdot \epsilon \log_2 \binom{M}{M/4} / \lceil \log_2 M \rceil$ , which grows to infinity as M increases. Hence, by choosing M sufficiently large, we have constructed an example that satisfies the proposition.

To prove this last proposition, we constructed an artificial decision rule where, with high probability, the communication cost is very low. However, Appendix A.1 presents a more natural decision rule (see Example 4) where the probability distribution over types is uniform and shows that even in this case, the communication cost of selfishness can be exponential for Bayesian implementation. Also, note that we used a uniform probability distribution in an auction setting for our example with ex post implementation. These cases show that the communication cost of selfishness can be severe in simple and common cases when we consider the average-case communication cost measure.

## 8.2 Decision Correspondence

Suppose we allow our decision rule to be relational rather than functional, giving in each state a non-empty set of acceptable outcomes. Then we have a decision correspondence, and for each equilibrium concept we have to consider two cases. On the one hand, if all selections of the decision correspondence are implementable in the concept, then our upper bounds are still valid because every protocol computing an acceptable outcome will compute an implementable selection of the decision correspondence. On the other hand, as was shown recently by [Fadel, 2005], if only some of the selections are implementable, then none of our upper bounds are maintained. This is because the selection that is computed with selfless agents might have an arbitrarily lower communication complexity than all the selections that are implementable. Even without taking into account incentive constraints (i.e., information sets and computation of transfers) in the new selection, the overhead is already unbounded. A fortiori, the communication cost of selfishness is unbounded in this case.

## 8.3 Dominant Strategy Implementation

**Definition 12.** BDM  $\langle \mathcal{P}, H, t \rangle$  is Dominant strategy Incentive Compatible (DIC) if in any state  $u \in U$ , the strategy  $s_i^* = \sigma_i(u_i)$  maximizes the utility of agent i, regardless of the strategies of the other agents:

$$\forall i \in I, \forall s \in S : u_i(x(g(s_i^*, s_{-i}))) + t_i(g(s_i^*, s_{-i})) \ge u_i(x(g(s))) + t_i(g(s)).$$

Since dominant strategy implementation is stronger than ex post, it is immediate that the average-case communication cost of selfishness is also unbounded. Furthermore, we have shown

in Section 5.2 that, as with Bayesian and ex post implementations, any simultaneous communication protocol that computes a DIC-implementable decision rule is DIC-incentivizable (even with interdependent valuations). Hence the exponential upper bound on the communication cost of selfishness holds with dominant strategy implementation, and can be proved along the same lines as the proof for ex post implementation.

Note that, contrary to ex post implementation, the restriction to perfect information is not without loss of generality for dominant strategy implementation. Intuitively, as in the Bayesian case, we need to hide information from the agents to reduce the set of available strategies. Also, as in the Bayesian case, the incentives of the agents can be maximized by using maximally large information sets. However, the reasons behind the need of reducing the number of deviations are different: in the Bayesian case, we need to reduce the number of deviations of an agent to satisfy the incentives of the agent himself, whereas in the dominant strategy case, we need to reduce the number of deviations of an agent to satisfy the incentives of the other agents.

## 8.4 Interdependent Valuations

With interdependent valuations, an outcome's valuation for some agent is determined not only by his type, but by the types of the other agents as well. In this case, the overhead is unbounded for expost implementation. We illustrate this result with the following example.

**Example 5.** We consider the incentives of only Agent 1, who has a bit  $b \in \{0,1\}$ . The desired outcome is the value of this bit b. Agent 1 receives no utility from outcome 0. Agent 2 is indifferent between the outcomes but his type determines the private valuation  $v_1 \in \{k2^{-\gamma} : k = 0, ..., 2^{\gamma} - 1\}$  of Agent 1 for outcome 1. The communication complexity of the rule is 1 bit (Agent 1 outputs b). Also, the decision rule is EPIC-implementable by paying Agent 1 transfer  $t_1(b, v_1) = (1 - b) \cdot v_1$ . However, the protocol needs to reveal the valuation  $v_1$  within  $2^{-\gamma}$  to enforce the ex post incentive constraints, which takes at least  $\gamma - 1$  bits. This can be made formal along the lines of the proof of Proposition 11. Hence the communication cost of selfishness is arbitrarily high in this case.

However, we still have an exponential upper bound with Bayesian implementation, as there is no overhead with simultaneous communication protocols. Intuitively, once the outcome has been computed, the mechanism can take the expectation over the possible transfers of each agent to satisfy their Bayesian incentives. Note that we need the crucial assumption that types be independently distributed.

### 8.5 Correlated Types

Our analysis of the overhead with ex post implementation need not be changed in this case, since our results do not depend on the space distribution. Furthermore, our results on the average-case communication cost of selfishness with both implementation are also not altered, as the overhead is already unbounded with independent types.

However, the communication cost of selfishness is unbounded for Bayesian implementation. We illustrate this result with the following example.

**Example 6.** We consider incentives of only Agent 1, who has a string w of M bits. The desired outcome is the parity of the string, i.e.,  $f(w) = \left(\sum_{m=1}^{M} w(m)\right) \mod 2$ . Agent 1 also gets a zero utility from outcome 0, and has utility that is either 1 or -1 for outcome 1, both with the same

probability 1/2. Agent 2's type is an integer m between 1 and M and the value of w(m). The distribution of m and w is uniform. The communication complexity of the rule is 1 bit (Agent 1 just outputs the outcome). Also, the direct revelation BDM satisfies BIC(p) with a high monetary punishment for Agent 1 "caught" lying, i.e., announcing a wrong value of w(m). However, any BIC(p)-incentivizable protocol must have depth at least  $\log_2 M$ , as otherwise there must be fewer than  $2^M$  strategies, and hence there would be 2 different types w and w' that share the same prescribed strategy in the protocol. Note that they must be of the same parity, say 0. But in this case, we could construct a type w'' that agrees on all the indexes where w and w' agree but which has parity 1. There would be no way for an agent having type w'' and preferring outcome 0 to be prevented from choosing the strategy of types w and w' (without preventing w or w' from being truthful). Hence the communication cost of selfishness is arbitrarily high in this case.

The above example shows that, without the independent types assumption, we cannot stop a BIC(p)-incentivizable protocol when the outcome is computed, and hence the computation of the transfers may cause an increase in the communication requirements.

# 9 Conclusion

We have examined the communication cost of selfishness in the independent private value settings. On the one hand, with ex post implementation, we have shown that the overhead comes only from the need compute a transfer rule that will satisfy the agents' incentives. On the other hand, with Bayesian implementation, we have shown that we never need additional information to compute transfers, and the overhead comes only from the need to hide information from the agents. Quantitatively, the communication cost of selfishness turns out to be at most exponential, and this upper bound is tight for the Bayesian implementation. Also, we have considered some special cases in which the communication cost of selfishness is low. These include in the expost setting the efficient decision rule case and the two agents case, as long as the utilities of the agents have some fixed precision. It also includes in the Bayesian setting the case where the decision rule is EPIC-implementable: the communication cost of selfishness in this case is zero. Finally, we have considered some extensions of our initial setting. In particular, we have shown that the communication cost of selfishness is unbounded when we allow interdependent valuations with expost implementation or correlated types with Bayesian implementation, or when we consider the average-case communication cost measure with either implementation concept.

The main open questions raised by the paper is the following:

- 1. How high is the communication cost of selfishness with expost implementation?
- 2. From a practical point of view, how broad are the cases in which the communication cost of selfishness is low?
- 3. Can the communication cost of selfishness be reduced substantially if agents' utilities have a given finite precision (or, equivalently, their incentive constraints need to be satisfied only approximately)?

We hope that these questions will be addressed in future research.

## A Technical Proofs

We begin with a simple lemma that is useful for bounding below the average-case communication complexity  $ACC_p(\mathcal{P})$  of a protocol  $\mathcal{P}$  (as defined in Subsection 8.1):

**Lemma 4.** If protocol  $\mathcal{P}$  has a subset L' of leaves whose aggregate probability is at least  $\alpha$  and the probability of each leaf from L' is at most  $\beta$ , then  $ACC_p(\mathcal{P}) \geq -\alpha \cdot \log_2 \beta$ 

Proof. We can consider the leaves L of  $\mathcal{P}$  as the possible realizations of a random variable  $\tilde{l}$ , each leaf  $l \in L$  having probability  $p(l) = \Pr(u \in U(l))$ . Shannon's theory of information ([Shannon, 1948], surveyed in [Cover and Thomas, 1991]) implies that  $ACC_p(\mathcal{P})$ , the average communication cost of  $\mathcal{P}$ , is bounded below by the entropy of  $\tilde{l}$ , defined as  $H(\tilde{l}) = -\sum_{l \in I} p(l) \log_2 p(l)$ .

In our case, the entropy of  $\tilde{l}$  is bounded below by the case in which a maximal number of leaves in L' have probability  $\beta$ . Let  $\alpha = n \cdot \beta + m$ , with  $n \in \mathbb{N}$  and  $0 \le m < \beta$ . Then

$$ACC_p(\mathcal{P}) \ge H(\tilde{l}) \ge -n\beta \log_2 \beta - m \log_2 m \ge -n\beta \log_2 \beta - m \log_2 \beta = -\alpha \cdot \log_2 \beta$$

## A.1 Bayesian Implementation: Exponential Overhead

Consider the following setting. The set of outcomes is  $X = \{x_1, \ldots, x_M\}$ . Agent 1's type is  $(u, \delta) \in U \times \Delta$ , where  $\Delta$  is the set of one-to-one correspondences over  $M = \{1, 2, \ldots, M\}$ , and where u describes Agent 1's utility function from  $U = [0, 1]^X$ . Agent 2's type is just one number  $m \in M$ , and he is always indifferent over the outcomes<sup>11</sup>. We consider the uniform state distribution p. The decision rule is  $f(\delta, u, m) = f(\delta, m) = x_{\delta(m)}$ .

The following protocol gives an upper bound on the communication complexity of the rule: Agent 2 outputs m (using  $\lceil \log_2 M \rceil$  bits) and then Agent 1 outputs  $\delta(m)$  (using  $\lceil \log_2 M \rceil$  bits). Hence:  $\mathrm{CC}(f) \leq 2 \lceil \log_2 M \rceil$ .

Also, the rule is BIC(p)-implementable: Consider the direct revelation mechanism without any transfer: Agent 1 with valuations u will receive expected utility  $\sum_{i=1}^{M} u(x_i)/M$  whatever is his announcement. Hence, all announcements will give the same expected payoff to Agent 1 and to Agent 2 (which is always indifferent), which implies that the rule is BIC(p).

Agent 2 (which is always indifferent), which implies that the rule is BIC(p). We now prove that  $CC_p^{BIC}(f) \ge 0.5 \log_2 \binom{M}{M/4}$  by proving the stronger statement  $ACC_p^{BIC}(f) \ge 0.5 \log_2 \binom{M}{M/4}$ . (Where  $ACC_p^{BIC}(f)$  is the average Bayesian incentive communication complexity, as defined in Section 8.1).

We will not count in the communication costs the information that Agent 1 gives about u during the execution of the protocol (but it must be finite). We will assume that, first, Agent 1 reports some information about his valuations that restricts u to some non-zero measure subset  $U' \subset U$ , and second, the BDM executes without asking questions to Agent 1 about his valuations. Note that since Agent 1 communicates more information earlier, this can only increase his incentives during the execution of the BDM, and also note that a lower bound on the communication costs during the execution of the BDM is a lower bound on the communication requirements of any BDM that might ask information about u during its execution.

<sup>&</sup>lt;sup>11</sup>Note that we allow both agents to have types that are not uniquely defined by their utility functions. This can be done formally by adding a fictitious outcome  $x_0$  for which each type gives a different "payoff."

Claim 1. If  $\mathcal{B}$  is a BDM that has finite average communication complexity and that implements f in BIC(p), then the updated distribution over m at any information set  $h \in H_1$  of Agent 1 with a strictly positive probability is the uniform distribution  $p^h$  over a set of indices  $M_h \subset M$  such that, for every legal permutation  $\delta$  at h,  $\delta(M_h)$  is the same fixed set  $M'_h$ .

*Proof.* We prove that the claim by contradiction. Suppose there is an information set h that does not satisfy the property. Then, consider the difference in expected value of the transfers for two distinct legal types  $(\delta, u)$  and  $(\delta', u)$  that disagree on the probability distribution on the outcome. The difference must be equal to the following linear combination of the exact valuations in  $u \in U'$ :

$$\Delta t = t_1 - t_1' = E[u(\delta'(m))] - E[u(\delta(m))] = \sum_{m \in M} p^h(m)[u(\delta'(m)) - u(\delta(m))]$$

Note that the above linear combination cannot be computed by the protocol as there are an infinite number of such values. Formally, for every difference  $\Delta t$ , the set of u for which the linear combination has this value must have zero measure. Since we could compute this difference in two runs of the protocol, and by Lemma 4 with  $\beta = 0$ , this would imply that the protocol has an infinite average communication cost. This cannot be the case, so the claim holds.

Claim 2. Any BDM that implements f in BIC(p) must, with probability 1/2, compute two sets  $M_1, M_2 \subset M$  such that:  $M/4 \leq \text{Card } M_1 = \text{Card } M_2 \leq M/2$  and  $\delta(M_1) = M_2$ .

Proof. Fix any permutation  $\delta$ . We can construct a tree  $T(\delta)$  (not necessarily binary) from the BDM by removing Agent 2's nodes and considering each of Agent 1's information sets  $h \in H_1$  as a single node labeled with  $M_h$ . The root will be labeled with M. A node n will be the parent of a node n', corresponding to information sets h and h' respectively, if and only if Agent 1 with permutation  $\delta$  can get from h to h' without going through any of his other information sets. We consider only information sets that have a strictly positive probability in the construction of the tree.

When Agent 1 has permutation  $\delta$ , we can trace the execution of the BDM through a path of this tree  $T(\delta)$ . Given  $\delta$ , for every execution (that is, for every m) that goes through a node n of  $T(\delta)$ , note that the protocol must have computed both the label  $M_{h'}$  and  $\delta(M_{h'})$  for every child n' of n in  $T(\delta)$ . This is because after making his move at h, the information set corresponding to n, Agent 1 must have described all the labels  $M_h$  of the children, along with  $\delta(M_h)$ , to be able to receive information about m from Agent 2 while still having all the children information sets satisfy the property of Claim 1. In other words, he cannot receive information about m before he describes the partitions of  $M_h$  and  $\delta(M_h)$  that is induced by the children nodes.

Let us walk from the root along some path of the tree  $T(\delta)$ , while always choosing the child that has the highest cardinality  $M_h$ . We continue until we get to a node n for which every child is labeled with some  $M_h$  that has cardinality less than M/2. First note that for at least M/2 of the m's (and hence with probability at least 1/2), the execution will go through n. Also note that, since all n's children's labels will be computed during every execution that goes through n, for each of these executions we can compute two sets  $M_1$  and  $M_2$  satisfying  $M/4 \leq \text{Card } M_1 = \text{Card } M_2 \leq M/2$  and  $\delta(M_1) = M_2$ . This is done by pooling the labels of some of the children nodes until we get a set  $M_1$  satisfying the above cardinality requirement. To conclude, note that we will compute these sets with probability 1/2 for every  $\delta$ , so this will hold independently of  $\delta$ .

Claim 3. Any protocol that computes, on 1/2 of the runs, two sets  $M_1$  and  $M_2$  such that:  $M/4 < \text{Card } M_1 = \text{Card } M_2 \leq M/2$  and  $\delta(M_1) = M_2$  must have an average communication cost that is at least  $0.5 \log_2 \binom{M}{M/4}$ .

Proof. Consider any of the leaves reached on these 1/2 "high cost" runs. The number of  $\delta$  that are consistent with some given  $M_1$  and  $M_2$  must be at most (3M/4)!(M/4)!. Since all the  $\delta$  are equiprobable, each of these leaves can only be reached with at most probability  $(3M/4)!(M/4)!/M! = 1/\binom{M}{M/4}$  for these "high cost" runs. Hence, by Lemma 4, the minimal average communication cost is at least  $0.5 \log_2 \binom{M}{M/4}$  on these runs.

Our proof does not count the information that Agent 1 might give about his valuations, and the lower bound extends to any BDM that implements f. So  $\mathrm{ACC}_p^{BIC}(f) \geq 0.5 \log_2 \binom{M}{M/4}$ , which implies that  $\mathrm{CC}_p^{BIC}(f) \geq 0.5 \log_2 \binom{M}{M/4}$ . This is exponentially higher than  $\mathrm{CC}(f) \leq 2\lceil \log_2 M \rceil$ .

## A.2 Ex Post Implementation: Unbounded Average-Case Overhead

Consider the problem of allocating an indivisible object to one of two agents, but with the agents' types drawn independently from a uniform distribution over  $U_1 = U_2 = U = \{k2^{-\gamma} : k = 0, ..., 2^{\gamma} - 1\}$ . Let f be the "efficient" decision rule that allocates the object to the agent with the higher valuation, and gives it to Agent 1 in the case of a tie. f can be computed using the following bisection protocol suggested in [Grigorieva et al., 2002]: At each round  $m = 1, ..., \gamma$ , each agent i reports the i the bit in the binary expansion of his valuation i. The protocol stops as soon as the two agent report different bits, and then the object is given to the agent who reported 1 (he is proven to have the higher valuation). If the agents have not disagreed after i steps, the object is allocated to Agent 1 (in this case the two agents are shown to have the same valuations, and either allocation would be efficient). At any given round, the probability that the protocol stops conditional on arriving there is i 1/2. Therefore, the expected number of rounds is at most 2, and so the average-case communication complexity is at most 4, regardless of i.

Now, consider an EPIC BDM that implements decision rule f (in fact, the argument below applies to any efficient decision rule). By (3), the transfer to Agent 2 can be written as  $\tau_2(f(u_1, u_2), u_1)$ . Furthemore, EPIC inequalities imply that

$$|\tau_2(2, u_1) - \tau_2(1, u_1) - u_1| \le 2^{-\gamma} \text{ for every } u_1 \in (0, 1 - 2^{-\gamma}),$$
 (4)

for otherwise Agent 2 would prefer either to understate his valuation when  $u_1 = u_2 - 2^{-\gamma}$  or to overstate it when  $u_1 = u_2 + 2^{-\gamma}$ .

Suppose that  $\gamma \geq 3$ . Let us now run the EPIC BDM twice with 3 agents whose valuations are drawn independently from U. The first run is with Agent 1 and Agent 2, and the second run is with Agent 1 and Agent 3 taking the place of Agent 2. Clearly the total average communication cost of the two runs is twice the average communication cost of the EPIC BDM.

In the event where Agent 2 has type  $u_2 \geq 3/4$ , Agent 3 has type  $u_3 < 1/4$ , and Agent 1 has type  $u_1 \in \tilde{U}_1 = \{k2^{-\gamma} : k = 1/4, \dots, 3/4 - 2^{-\gamma}\}$  (this event that has probability  $1/4 \cdot 1/4 \cdot 1/2 = 1/32$ ), the object goes to Agent 2 in the first run and to Agent 1 in the second run, and, by (4) the difference between Agent 2's and Agent 3's transfers pins down the realization of  $u_1 \in \tilde{U}_1$  within  $2^{-\gamma}$ . Thus each of the pair of runs cannot have a probability more than  $3 \cdot 2^{-\gamma}$ . Hence by Lemma 4, the average communication complexity of the two runs is at least  $1/32 \cdot \log_2(2^{\gamma}/3) > (\gamma - 2)/32$ . The average-case communication cost of a single run of the EPIC BDM is then at least half this number, i.e.,  $(\gamma - 2)/64$ . We can then choose  $\gamma$  to get an arbitrarily high average-case communication cost.

# References

- [Athey and Segal, 2005] Susan Athey and Ilya Segal. An efficient dynamic mechanism, 2005. Working paper.
- [Cover and Thomas, 1991] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., N. Y., 1991.
- [Fadel and Segal, 2005] Ronald Fadel and Ilya Segal. The communication cost of selfishness: Ex post implementation. In *Proc. 10th Conference in Theoretical Aspects of Rationality and Knowledge*, 2005.
- [Fadel, 2005] Ronald Fadel. A contribution to canonical hardness, 2005. The Seventh International Workshop on Agent-Mediated Electronic Commerce (AMEC 2005).
- [Feigenbaum et al., 2002] J. Feigenbaum, A. Krishnamurthy, R. Sami, and S. Shenker. Hardness results for multicast cost sharing, 2002. Yale University Technical Report YALEU/DCS/TR1232, June 2002.
- [Forges, 1990] Francoise Forges. Equilibria with communication in a job market example. Quarterly Journal of Economics, 105:375–398, 1990.
- [Fudenberg and Tirole, 1991] Drew Fudenberg and Jean Tirole. Game Theory. MIT Press, 1991.
- [Green and Laffont, 1987] Jerry Green and Jean-Jacques Laffont. Limited communication and incentive compatibility. In *Information, Incentives, and Economic Mechanisms: Essays in honor of Leonid Hurwicz*, 1987.
- [Grigorieva et al., 2002] Elena Grigorieva, P. Jean-Jacques Herings, Rudolf Muller, and Dries Vermeulen. The private value single item bisection auction. Research Memoranda 051, Maastricht: METEOR, Maastricht Research School of Economics of Technology and Organization, 2002.
- [Johari, 2004] R. Johari. Efficiency loss in market mechanisms for resource allocation. PhD thesis, Massachusetts Institute of Technology, 2004.
- [Kushilevitz and Nisan, 1997] E. Kushilevitz and N. Nisan. Communication Complexity. Cambridge University Press, 1997.
- [L. Blumrosen and Segal, ] N. Nisan L. Blumrosen and I. Segal. Multi-player and multi-round auctions with severely bounded communication. ESA 2005.
- [Lahaie and Parkes, 2004] S. Lahaie and D. Parkes. Applying learning algorithms to preference elicitation, 2004.
- [Lavi et al., 2003] R. Lavi, A. Mu'alem, and N. Nisan. Towards a characterization of truthful combinatorial auctions. In Proc. of the 44th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2003.
- [Mas-Colell et al., 1995] Andreu Mas-Colell, Michael Whinston, and Jerry R. Green. Microeconomic Theory. 1995.
- [Melumad et al., 1992] N. Melumad, D. Mookherjee, and S. Reichelstein. A theory of responsibility centers. Journal of Accounting and Economics, 15:445–484, 1992.

- [Mookherjee, 2006] D. Mookherjee. Decentralization, hierarchies and incentives: A mechanism design perspective. *Journal of Economic Literature*, 2006. forthcoming.
- [Myerson, 1991] Roger B. Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, Cambridge Mass., 1991.
- [Nisan and Segal, 2004] N. Nisan and I. Segal. The communication requirements of efficient allocations and supporting prices, 2004. Forthcoming in the Journal of Economic Theory.
- [Reichelstein, 1984] S. Reichelstein. Incentive compatibility and informational requirements. *Journal of Economic Theory*, 1984.
- [Segal, 2005] Ilya Segal. Communication in economic mechanisms. In Advances in Economic Theory: 9th World Congress. Blackwell, 2005.
- [Shannon, 1948] C. E. Shannon. A mathematical theory of communication. Bell System Tech. J., 27:379–423, 623–656, 1948.
- [Yao, 1979] Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In STOC '79: Proceedings of the eleventh annual ACM symposium on Theory of computing, pages 209–213. ACM Press, 1979.