Available online at www.sciencedirect.com
ScienceDirect
Journal of Economic Theory 144 (2009) 1895–1920
JOURNAL OF Economic Theory
www.elsevier.com/locate/jet

The communication cost of selfishness

Ronald Fadel [a], Ilya Segal [b,*]

[a] *ABC Group, Lebanon*
[b] *Department of Economics, Stanford University, Stanford, CA 94305, United States*

Received 26 January 2006; final version received 10 April 2007; accepted 13 September 2007
Available online 23 February 2009

Abstract

We consider how many bits need to be exchanged to implement a given decision rule when the mechanism must be ex post or Bayesian incentive compatible. For ex post incentive compatibility, the communication protocol must reveal enough information to calculate monetary transfers to the agents to motivate them to be truthful (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 settings with selfish agents, the communication cost can be higher than in the case in which the agents are honest and can be relied upon to report truthfully. The increase is the "communication cost of selfishness." We provide an exponential upper bound on the increase. We show that the bound is tight in the Bayesian setting, but we do not know this in the ex post setting. We describe some cases where the communication cost of selfishness proves to be very low.

JEL classification: D82; D83

Keywords: Communication complexity; Algorithmic mechanism design; Bayesian incentive compatibility; Ex post incentive compatibility; Sequential and simultaneous communication protocols; Information sets

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

* Corresponding author.
E-mail address: ilya.segal@stanford.edu (I. Segal).

0022-0531/$ – see front matter
doi:10.1016/j.jet.2007.09.015

describe their preferences, and checks their incentives to do so truthfully (e.g., [17, 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 the value of each of the $2^L - 1$ non-empty 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., [15,21,24], and references therein). However, in most practical settings we should expect agents to communicate strategically to maximize their own benefit.

This paper considers how many bits need to be exchanged among agents to implement a given decision rule when the agents are selfish, and compares it to the number of bits required when they are honest (the latter is known as "communication complexity"). We will refer to the difference as the *communication cost of selfishness*, or, for short, the *overhead*. In cases in which the overhead is high, economic goals that could be achieved with selfish agents but extensive communication or with honest agents but limited communication are not achievable when the agents are selfish and communication is limited at the same time.[1]

For a simple illustration, consider the problem of allocating an indivisible object efficiently between two agents, who have privately known values for the object (which is formally described in Example 1). In a direct revelation mechanism, we can make truthtelling a dominant strategy for each agent by having the winner of the object pay a price equal to the loser's report (as in the second-price sealed-bid auction). We can try to reduce the communication cost by first asking Agent 1 report his value and then asking Agent 2 to say whether his valuation exceeds Agent 1's report. (Since Agent 2 sends only one bit, the communication cost is reduced from full revelation roughly by half if the two agents' values have the same range, and possibly by more if they have different ranges.) The problem with the new communication protocol, however, is that it does not reveal enough information to construct payments to Agent 1 that make truthtelling a dominant strategy or even an ex post best response for him. Intuitively, any such payments must face Agent 1 with a price approximately equal to Agent 2's valuation, but this valuation is not revealed. We show that this mechanism cannot be incentivized with any payments, and that any ex post incentive compatible mechanism that computes the efficient allocation in this example requires sending more bits. Thus, the communication cost of ex post incentive compatibility is strictly positive.

In general, a mechanism designer can use two instruments to motivate agents to be honest: First, along with computing the desired allocation, she could use the communication protocol to compute transfers to the agents (as in the above example). 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.

This paper analyzes the communication cost of selfishness for two distinct equilibrium concepts: *Ex Post Incentive Compatibility* (*EPIC*), in which an agent should report honestly even

[1] Of course, prohibitive communication within a mechanism is not the only possible reason for why it may be impractical. For example, a mechanism may be difficult to explain to the agents, even if, once they understand it, it is easy to execute. However, in situations in which the agents face the same type of problem and can use the same mechanism again and again, the "fixed cost" of explaining the mechanism to the agents is amortized over a long horizon, and the practicality of the mechanism is determined by the "variable cost" of communication within it.

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 the other agents' information. In both settings, we focus on the case in which agents have private and independently drawn valuations over outcomes, their utilities are quasilinear in monetary transfers, and the communication cost is defined as the maximal number of bits sent during the execution of the mechanism.

For both EPIC and BIC implementation, we show that the communication cost of selfishness may be strictly positive. However, the reasons for the overhead differ between the two cases: For EPIC, there is no need to hide information from agents, and the overhead comes entirely from the need to compute motivating transfers. For BIC, 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 computing an implementable decision rule can be incentivized with some transfers. Intuitively, in such a protocol, we fully observe all agents' strategies, and this proves sufficient to compute incentivizing transfers. However, minimizing communication cost typically requires sequential (extensive-form) communication (so that the information reported at a given stage is contingent on what has been reported previously), and such communication does not fully reveal the agents' contingent strategies.

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 resulting simultaneous protocol can then be incentivized. If the original sequential protocol communicated no more than b bits, then the number of bits announced 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, and several examples in which it is fairly low. For example, the EPIC overhead is low for efficient allocation rules (such as the example above), for which we can simply ask the agents to report their utilities at the end of a protocol, and give each agent the sum of the other agents' reports. On the other hand, for the BIC case, we do provide an example with an exponential overhead. This example (formally described in Example 2 below) can be interpreted as having an expert with private knowledge and a private utility function, and a manager with a private goal that determines how the expert's knowledge is used. The expert will reveal his knowledge truthfully if he does not know the manager's goal, but this revelation requires long communication. Communication cost could be reduced exponentially by having first the manager announce his goal and then the expert say 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 exponentially longer — almost as long as full revelation of the expert's knowledge.

This example notwithstanding, we do find several cases in which the communication cost of Bayesian incentives is low. In particular, it is zero for any EPIC-implementable rule (including, e.g., efficient decision rules, such as in the allocation problem described above). Namely, we show that any communication protocol that computes such a rule can be BIC incentivized by computing transfers on the basis of its outcome, and without hiding any information from the agents.

Finally, in Appendices A and B we consider several extensions of the model. In particular, we consider the communication cost measured as the *average-case* rather than worst-case number of bits sent, given a probability distribution over agents' private information. We also consider cases in which the agents' valuations are interdependent or correlated. In all these cases, we provide examples in which the communication cost of selfishness can grow without bound even as the communication cost with honest agents is fixed at just a few bits.

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 [22], who considered incentive compatibility in non-deterministic real-valued mechanisms, and showed that the communication cost of selfishness in achieving efficiency is low. Lahaie and Parkes [16] characterized the communication problem of finding Vickrey–Clarke–Groves (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. [9].

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, they ask 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., [12,18], and the recent survey [19]). A similar question is studied in [14], which instead focuses on the efficiency objective.

Finally, the literature on communication without commitment ("cheap talk") has offered examples in which incentive-compatible communication requires a large number of stages (e.g., [10]). 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 (yet 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 [26] and surveyed in [15], describes how many bits agents from a set $I = \{1, \dots, I\}$ must exchange in order to compute the value of a function $f : \prod_{i \in I} U_i \to X$ when each agent $i \in I$ knows privately its argument $u_i \in U_i$, which we refer to as agent i's "type."[2] 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,

[2] With a slight abuse of notation, we will use the same letter to refer to a set and its cardinality.

Definition 1. A protocol $\mathcal{P}$ with the set of agents I over state space $U = \prod_{i \in I} U_i$ and outcome space X is a tuple $\langle N_1, \ldots, N_I, L, r, c_0, c_1, x, \sigma_1, \ldots, \sigma_I \rangle$, where:

- The sets $N_1, \ldots, N_I$ and L are pairwise disjoint. N_i represents the set of decision nodes of agent $i \in I$, while L denotes the set of terminal nodes (leaves). Let $N = (\bigcup_{i \in I} N_i) \cup L$ — the set of all nodes of the protocol.
- The protocol forms a binary tree with root $r \in N$ and the child relation described by $c_0, c_1 : N \backslash L \to N \backslash \{r\}$. That is, $c_0(n)$ and $c_1(n)$ represent the two children of node $n \in N \backslash L$. Each node $n \in N \backslash \{r\}$ has a unique "parent" n' such that either $n = c_0(n')$ or $n = c_1(n')$.
- $x : L \to X$, where $x(l) \in X$ is the outcome implemented at leaf $l \in L$.
- For each agent $i \in I$, $\sigma_i : U_i \to \{0, 1\}^{N_i}$ is the agent's strategy plan in the protocol, where $\sigma_i(u_i) \in \{0, 1\}^{N_i}$ specifies the strategy of the agent of type u_i — the moves he makes at each of his decision nodes. That is, at each decision node $n \in N_i$, the agent moves to node $c_{\sigma_i(u_i)(n)}(n)$.[3]

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 $\mathcal{P}$ is defined by $f = x \circ g \circ \sigma$, and denoted by $Fun(\mathcal{P})$.

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 states in which node n is reached. For example, at the protocol's root r, $U(r) = U$. 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 possible number of moves between the root and a leaf — i.e., the number of bits sent in the protocol in the worst case.[4] The communication complexity of a function is the minimal depth of a protocol computing it:

Definition 2. The communication complexity of a function $f : U \to X$ is $\mathrm{CC}(f) = \inf_{\mathcal{P}:\, Fun(\mathcal{P}) = f} d(\mathcal{P})$.

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 his type. With a slight abuse of notation, we identify each agent i's type

[3] It is customary in game theory to call the "strategy" of agent i the whole function σ_i, by interpreting the agent's type $u_i \in U_i$ 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.

[4] We consider average-case communication costs in Appendix B.1.

with his utility function $u_i : X \to \mathbb{R}$ over a set X of possible outcomes.[5] We assume that the agents' payoffs are quasilinear in monetary transfers, hence the total payoff of agent i of type $u_i \in U_i$ from outcome $x \in X$ and transfer t_i is $u_i(x) + t_i$. In particular, with such utilities, we will often be concerned with implementing a decision rule f that is *efficient*, i.e., satisfies $f(u) \in \arg\max_{x \in X} \sum_{i \in I} u_i(x), \forall u \in U$.

A protocol induces an extensive-form game, and prescribes 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.[6] 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 $\langle \mathcal{P}, H, t \rangle$, where

- $\mathcal{P}$ is a protocol $\langle N_1, \ldots, N_I, L, r, c_0, c_1, x, \sigma_1, \ldots, \sigma_I \rangle$.
- $H = (H_1, \ldots, H_I)$, where H_i is the information partition of agent i — a partition of his decision nodes N_i into information sets satisfying perfect recall.[7]
- $t = (t_1, \ldots, t_I)$, where $t_i : L \to \mathbb{R}$ is the transfer function for agent i. Thus, $t_i(l)$ is the transfer made to agent i when leaf $l \in L$ is reached.
- The agents' strategies in protocol $\mathcal{P}$ are consistent with their information partitions. Namely, say that agent i's strategy $s_i \in \{0,1\}^{N_i}$ is consistent with his information partition if it is constant on any element of H_i, and let $S_i \subset \{0,1\}^{N_i}$ denote the set of such strategies. We require that for every agent $i \in I$ and type $u_i \in U_i$, $\sigma_i(u_i) \in S_i$.

Note that when the information partition of agent i becomes coarser (and hence its cardinality $|H_i|$ is reduced), his strategy space shrinks. The communication cost (or just depth) $d(\mathcal{B})$ of a BDM $\mathcal{B} = \langle \mathcal{P}, H, t \rangle$ is the depth $d(\mathcal{P})$ of its protocol $\mathcal{P}$.

4.2. Two particular classes of BDMs

It is useful to define two extreme cases of information revelation to the agents in mechanisms. On the one hand, we define a *Perfect information BDM* (PBDM) as a BDM $\langle \mathcal{P}, H, t \rangle$ in which the agents observe the complete history of messages — i.e., every information set $h \in H$ is a singleton. On the other hand, we define a *Simultaneous communication BDM* (SBDM) as a BDM in which agents never learn anything other than his previous moves, and so the game is strategically equivalent to one in which all agents send all messages simultaneously. Formally, an SBDM is a BDM $\langle \mathcal{P}, H, t \rangle$ in which any two decision nodes n and n' of an agent i that have the same history of moves for agent i are in the same information set.

[5] Sometimes we may want to allow different types to have the same utility function over outcomes from X, which can be done by adding a fictitious outcome that gives a different utility for each of the types.

Note also that this formalism assumes *private values*, i.e., that the utility function of an agent is determined by his own type. This assumption will be relaxed in Appendix B.3.

[6] Our restriction to agents following the prescribed strategies is without loss of generality, for we can prescribe any possible strategy profile of the game.

[7] 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 [11, p. 81] for more details.

In particular, a *direct revelation BDM* is an SBDM in which the mapping $g \circ \sigma$ from the state space into leaves is one-to-one (note that this is only possible when the state space is finite). That is, at the end of the execution of the BDM the state is completely revealed but no agent has observed anything besides his own moves.

Note that some protocols cannot be used in an SBDM, since their strategy plan may not be compatible with simultaneous communication. A *simultaneous communication protocol* is defined as a protocol that can be part of an SBDM. The following lemma will prove useful:

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

Proof. Construct protocol $\mathcal{P}'$ by having each agent report his strategy in $\mathcal{P}$, and prescribe the outcome that would obtain in $\mathcal{P}$ given the reported strategies. (Since we require that all nodes have a non-empty domain, the agents should be restricted to report only strategies in $\mathcal{P}$ that are used by some types.) Since each agent's strategy in $\mathcal{P}'$ is not contingent on the other agents' moves, $\mathcal{P}'$ is a simultaneous communication protocol. To count the number of possible strategies in $\mathcal{P}$, note that each agent i will output at most one bit for each of his decision nodes $n \in N_i$ in $\mathcal{P}$, so the depth of $\mathcal{P}'$ is at most $\sum_{i \in I} |N_i| = |N \setminus L| \leqslant 2^{d(\mathcal{P})} - 1$. $\square$

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 $s = (\sigma_1(u_1), \dots, \sigma_I(u_I)) \in \prod_{i \in I} S_i$ is an ex post Nash equilibrium of the induced game, i.e.,

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

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.[8] 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:

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

$$U_{-i}(l) \cap U_{-i}(l') \neq \emptyset \quad \Rightarrow \quad \forall u_i \in U_i(l): \quad u_i\big(x(l)\big) + t_i(l) \geqslant u_i\big(x(l')\big) + t_i(l'). \tag{1}$$

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 $\{l' \in L\colon\ U_{-i}(l) \cap U_{-i}(l') \neq \emptyset\}$ by deviating, as it is the set of leaves

[8] In our setting of *private values*, EPIC 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 Appendix B.2.

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 $U(l)$. $\square$

It immediately follows from Lemma 2 that the information partition is 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 partitions, $\mathcal{B}$ is EPIC if and only if $\mathcal{B}'$ is EPIC.*

Hence, we can restrict attention without loss to PBDMs when we are concerned with EPIC.

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:

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

Hence, all the EPIC-implementable rules we will consider in our private valuations setting are necessarily DIC-implementable. In particular, using (2) 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

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

4.4. Bayesian incentive-compatibility

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

Definition 5. Given a 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

$$\begin{aligned}&\forall i \in I,\ \forall u_i \in U,\ \forall s_i' \in S_i:\quad \mathbb{E}_{u_{-i}}\big[u_i\big(x\big(g\big(\sigma(u)\big)\big)\big) + t_i\big(g\big(\sigma(u)\big)\big)\big|u_i\big]\\ &\quad \geqslant \mathbb{E}_{u_{-i}}\big[u_i\big(x\big(g\big(s_i', \sigma_{-i}(u_{-i})\big)\big)\big) + t_i\big(g\big(s_i', \sigma_{-i}(u_{-i})\big)\big)\big|u_i\big].\end{aligned}$$

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 strategy plans. In the main text of the paper, we focus on the case where the types of the agents are independently distributed, i.e., the probability distribution p over states is a product of the individual probability distributions p_i over each agent i's type, and so the expectations above need not be conditioned on u_i. (This assumption will be relaxed in Appendix B.4.)

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

4.5. Incentivizability

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 an information partition H of $N \setminus L$ such that $\langle \mathcal{P}, H, t \rangle$ is an EPIC BDM.

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 an information partition H of $N \setminus L$ such that $\langle \mathcal{P}, H, t \rangle$ is a BIC(p) BDM.

Consider first EPIC incentivizability. By Lemma 2, it is equivalent to checking that there exists a transfer rule that solves the system of linear inequalities, which can be done with standard linear programming techniques (and regarding the information partition, we can restrict attention to PBDMs). Note that by definition, a protocol is EPIC-incentivizable only if it computes an EPIC-implementable decision rule. However, the converse is not true: not every protocol computing an EPIC-implementable decision rule is EPIC-incentivizable:

Example 1. There are two agents and one indivisible object, which can be allocated to either agent (and so we can write $X = \{(1, 0), (0, 1)\}$). The two agents' values (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). The efficient decision rule f is EPIC-implementable (e.g., using the Vickrey transfer rule $t_i(u_i, u_{-i}) = -f_i(u_i, u_{-i})u_{-i}$ for each i). The efficient decision rule 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 allocation $f(u_1, u_2)$ (using 1 bit). Suppose in negation that protocol $\mathcal{P}_0$ computes a transfer rule $t_1 : U_1 \times U_2 \to \mathbb{R}$ that satisfies the ex post incentives of Agent 1. Given the information revealed in the protocol, $t_1(u_1, u_2)$ can depend only on u_1 and $f(u_1, u_2)$, so it can be written as $t_1(u_1, u_2) = t_1'(u_1, f(u_1, u_2))$ for some $t_1' : U_1 \times X \to \mathbb{R}$. However, by (3), EPIC requires that t_1 satisfy $t_1(u_1, u_2) = \tau_1(f(u_1, u_2), u_2)$ for some $\tau_1 : X \times U_2 \to \mathbb{R}$. Hence $t_1(u_1, u_2) = t_1^*(f(u_1, u_2))$ for some $t_1^* : X \to \mathbb{R}$. But then if $t_1^*(1, 0) - t_1^*(0, 1) \leqslant 2.5$, 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 $t_1^*(1, 0) - t_1^*(0, 1) > 2.5$, 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 incentivizable, hence the communication cost of selfishness is positive.[9]

[9] 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

Now we turn to BIC incentivizability. By definition, a protocol is BIC(p)-incentivizable only if it computes a BIC(p)-implementable decision rule. However, the converse is not true: not every protocol computing a BIC(p)-implementable decision rule is BIC(p)-incentivizable:

Example 2. There are two agents — Agent 1 is the Expert and Agent 2 is the Manager. In addition to the set of outcomes X, there is a set of consequences M, with $|M| = |X| = k$. The Manager's private type is his "desired" consequence $m \in M$. The Expert's private type is a pair (u, δ), where $u : X \to [0, 1]$ is his utility function over outcomes, and $\delta : M \to X$ is a one-to-one mapping, where $\delta(m)$ specifies the outcome needed to achieve consequence m.[10] For simplicity, we abstract from the Manager's incentives by assuming that his utility over outcomes is identically zero. The decision rule f implements the outcome that yields the Manager's desired consequence m in to the mapping δ, i.e., $f(m, (u, \delta)) = \delta(m)$. The two agents' types are independently and uniformly distributed on their respective domains.

We assume that the realized consequence is not contractible (e.g., it is not observed by the Manager until after the mechanism is finished), so the transfers cannot be based on it, but only on the agents' messages. 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 regardless of the mapping δ he reports. Hence the decision rule is BIC(p)-implementable. With an honest Expert, the outcome can instead be computed with the following much simpler protocol $\mathcal{P}_0$: the Manager reports the consequence $m \in M$ he wants, and the Expert reports the corresponding decision $\delta(m)$. However, if the Expert is selfish, he will always report an outcome that maximizes his utility plus the transfer, and so no transfer rule computed by $\mathcal{P}_0$ can induce him to be truthful for all possible utility functions u. Hence, $\mathcal{P}_0$ is not BIC(p)-incentivizable. In fact, we will show in Section 6.3 that any BIC(p) mechanism must take exponentially more communication than $\mathcal{P}_0$ and a comparable amount to a direct revelation mechanism.

4.6. *Incentive communication complexity*

As we just saw, the cheapest communication protocol that computes a given EPIC (BIC(p))-implementable decision rule may not be EPIC (respectively, BIC(p))-incentivizable. So the need to satisfy incentive constraints may require an increase in the communication cost:

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

The communication cost of selfishness (*overhead* for short) can now be defined as the difference between $\mathrm{CC}(f)$ and $\mathrm{CC}^{EPIC}(f)$ in the ex post setting, or the difference between $\mathrm{CC}(f)$ and $\mathrm{CC}_p^{BIC}(f)$ in the Bayesian setting with state distribution p.

Agent 1 announces an efficient decision. In this protocol, the EPIC constraints of Agent 2 cannot be satisfied with any transfer, by an argument similar to that in Example 1.

[10] Even though δ does not affect the agent's utility over X, this model still fits in our framework as explained in footnote 5 above.

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, and so a restriction to 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 comes only from the need to reveal 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 ex post incentive constraints (2) 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, $\{f_t\colon\ t \in T\}$ is the set of EPIC social choice rules implementing f, and $\mathrm{CC}^{EPIC}(f)$ is exactly $\min_{t\in T} \mathrm{CC}(f_t)$. 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. An exponential upper bound

We now show that the communication cost of selfishness in EPIC implementation can be bounded by an (exponential) function of the communication complexity of the decision rule. We do this by noting that for a given EPIC-implementable decision rule, full revelation of types, while sufficient, may not be necessary to find incentivizing transfers. In fact, we can show that if we are given a protocol that computes an implementable decision rule, it is sufficient to reveal what each agent would do for all possible messages of the other agents. (That is, it suffices to reveal each agent's strategy in the protocol, with two strategies viewed as equivalent whenever they prescribe the same move for all possible messages of the other agents.) In particular, in an SBDM, an agent's strategy is not conditioned on the other agents' messages, and so we can state:

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

Proof. A simultaneous communication protocol $\mathcal{P}$ can be thought of as encompassing I single-agent protocols: For each agent i, let L_i be the set of possible move sequences by agent i, and let $g_i(s_i) \in L_i$ be the agent's move sequence when he uses strategy $s_i \in S_i$. The set of leaves of $\mathcal{P}$ can then be described as $L = \prod_{i\in I} L_i$, and the outcome function as $g(s) = (g_1(s_1), \dots, g_I(s_I))$. The legal domain of leaf $l = (l_1, \dots, l_I)$ takes the form

$$U(l) = \prod_{i\in I} U_i(l_i), \quad \text{where } U_i(l_i) = \{u_i \in U_i\colon\ g_i(\sigma_i(u_i)) = l_i\}. \tag{4}$$

Now, for each agent i, fix a selection $\gamma_i : L_i \to U_i$ such that $\gamma_i(l_i) \in U_i(l_i)$ for all $l_i \in L_i$. Let $\gamma = (\gamma_1, \dots, \gamma_I)$. Pick a transfer rule $t : U \to \mathbb{R}^I$ that satisfies the incentive constraints (2), and define $\bar{t} : L \to \mathbb{R}^I$ by $\bar{t} = t \circ \gamma$.

We now argue that the SBDM $\langle \mathcal{P}, H, \bar{t}\rangle$ (with any information partition H) is EPIC, using the characterization in Lemma 2. For this purpose, observe that under (4) the inequalities (1) amount to saying that for any $l \in L$, $l_i' \in L_i$, and any $u_i \in U_i(l_i)$, we must have

$$u_i\big(x(l)\big) + \bar{t}_i(l) \geqslant u_i\big(x\big(l_i', l_{-i}\big)\big) + \bar{t}_i\big(l_i', l_{-i}\big).$$

Let $u_{-i} = \gamma_{-i}(l_{-i})$ and $u_i' = \gamma_i(l_i')$. Since $\mathcal{P}$ implements f, we have $x(l) = f(\gamma_i(l_i), u_{-i}) = f(u_i, u_{-i})$ and $x(l_i', l_{-i}) = f(u_i', u_{-i})$. By (3) (which is implied by (2)), we must also have $\bar{t}_i(l) = t_i(\gamma_i(l_i), u_{-i}) = t_i(u_i, u_{-i})$. Thus, the above inequality follows from (2). □

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.

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

Proof. For any protocol $\mathcal{P}$ that achieves the lower bound $\mathrm{CC}(f)$, by Lemma 1, there is a simultaneous communication protocol $\mathcal{P}'$ computing the same decision rule such that $d(\mathcal{P}') \leqslant 2^{d(\mathcal{P})} - 1 = 2^{\mathrm{CC}(f)} - 1$. 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 implies that any EPIC implementable decision rule f that can be computed with finite communication can also be EPIC implemented in a finite BDM (even if the state space U is infinite).[11]

The upper bound of Corollary 2 can be improved upon by eliminating those strategies in the simultaneous protocol that are not used by any type in the original protocol.

Example 3. Consider the efficient object allocation setting described in Example 1. Protocol $\mathcal{P}_0$ has depth 3, so by Corollary 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 object. 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 $\mathrm{CC}^{EPIC}(f)$ is even close to this exponential bound of $2^{\mathrm{CC}(f)}$. In particular, an open problem is to determine the highest attainable upper bound, and to determine if there are any instances in which the incentive communication complexity of decision rule f, $\mathrm{CC}^{EPIC}(f)$, is much higher than the communication complexity of f, $\mathrm{CC}(f)$.

5.3. Low overhead for efficient decision rules

It is well known that any efficient decision rule is EPIC-implementable (and implementable in dominant strategies) by giving each agent a payment equal to the sum of the 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

[11] Our restriction to private values and our use of the worst-case communication cost measure are both crucial for this result. With interdependent valuations, or with the average-case communication cost, the EPIC overhead can be arbitrarily high, as we show in Appendices B.3, B.1 respectively.

a payment equal to the sum of the other agents's reported utilities. This approach dates back to Reichelstein [22] and was more recently used in [2], where it is called "the Team Mechanism." This approach has a low communication overhead when the utility ranges of the agents have low cardinality. For simplicity, we consider the case when agents' utilities are given in discrete multiples:

Definition 8. The utility function space U has discrete range with precision γ if $u_i(x) \in \{k2^{-\gamma}: k = 0, \dots, 2^\gamma - 1\}$ for all $x \in X$, $i \in I$, $u_i \in U_i$.

In this case, we can modify any efficient protocol to make it EPIC-incentivizable as follows: At each leaf l, let each agent report his utility $v_i = u_i(x)$ using γ bits, and give each agent a transfer $t_i = \sum_{j \neq i} v_j$. Thus, we have

Proposition 2. *For a utility function space U with discrete precision-γ range, and an efficient decision rule f,*

$$\mathrm{CC}^{EPIC}(f) \leqslant \mathrm{CC}(f) + I\gamma.$$

Thus, the communication cost of selfishness is at most γI bits.

Thus, the EPIC overhead of efficiency is bounded above by a number that does not depend on the communication complexity of the decision rule, and it is low if the utility range has a low cardinality.[12] Other low-overhead cases are considered in Appendix A.

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 a maximally coarse information partition, subject to perfect recall and to giving each agent enough information to compute his prescribed move at each step. (E.g., the information partition in an SBDM is maximally coarse because each agent never observes the others' moves during the execution of the protocol.) However, in general even the maximally coarse information partition may not ensure Bayesian incentives. In such cases, the need to hide information from the agents will require more information to be communicated simultaneously rather than sequentially, which may create a substantial BIC overhead.[13]

In contrast, computing incentivizing *transfers* is not a problem in itself with Bayesian implementation, as the following useful result suggests:

[12] The result extends easily to decision rules f that are "affine maximizers" [4], which can be interpreted as efficient upon rescaling the agents' utilities and adding a fictitious agent with a known utility. [23] shows that all decision rules that are DIC implementable on the universal domain are affine maximizers.

[13] The idea of hiding as much information as possible to maximize incentives can be traced back to Myerson's "communication equilibrium" [20] 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 minimizes the set of his feasible deviations.

Lemma 3. *Let $\mathcal{P}$ be a BIC(p)-incentivizable protocol computing decision rule f, for some product state distribution p. Then the protocol $\mathcal{P}'$ obtained from $\mathcal{P}$ by stopping its execution as soon as the outcome is known 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 suppose the left and right children of some node n are leaves l and l' that satisfy $x(l) = x(l') = x$, then the protocol $\mathcal{P}'$ obtained from $\mathcal{P}$ by removing l, l' and making n a leaf with outcome $x(n) = x$ is also BIC(p)-incentivizable. To prove this statement, 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 a 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 focus on satisfying the incentives of agent i. If $h = \{n\}$, agent i's incentives in the original BDM require that $t_i(l) = t_i(l')$ (because both leaves have a non-empty domain) and so 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 exactly the same for him to be at n, l or l' in $\mathcal{B}$. If instead 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 by 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 leaf 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 incentive constraints of agent i are satisfied in $\mathcal{B}'$. □

The lemma means that if we have a BIC(p) BDM and if at some node n of this BDM we have enough information to compute the outcome, we also have enough information to compute an incentivizing transfer $t(n)$, and so we do not need any more communication to satisfy the agents' incentive constraints.[14]

In general, it is hard to check if a protocol is BIC(p)-incentivizable.[15] However, a simple sufficient condition for a BDM to be BIC(p) is that no information that an agent i could possibly receive during the execution of the BDM (whatever strategy $s_i \in S_i$ he uses) could ever make his prescribed strategy suboptimal. Formally:

Proposition 3. *A protocol $\mathcal{P}$ computing decision rule f is BIC(p)-incentivizable with information partition H for some product distribution p if there is a transfer rule $t : U \to \mathbb{R}^I$ such that, for every agent $i \in I$ and each his information set $h \in H_i$ that is reached with a positive probability, 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.,*[16]

$$\forall u_i, u'_i \in U_i,\ \mathbb{E}_{u_{-i}}\big[u_i\big(f(u)\big) + t_i(u)\big|h\big] \geqslant \mathbb{E}_{u_{-i}}\big[u_i\big(f\big(u'_i, u_{-i}\big)\big) + t_i\big(u'_i, u_{-i}\big)\big|h\big].$$

Proof. Let $\bar{t}_i(l) = \mathbb{E}_u[t_i(u)|u \in U(l)]$, where the expectation uses distribution p. Then the BDM $\langle \mathcal{P}, H, \bar{t}\rangle$ is BIC(p). Indeed, for any possible deviation, and whatever information he learns about

[14] The assumed independence of types is crucial for this result. We will consider correlated types in Appendix B.4.

[15] Verifying that a given BDM is BIC(p) could be done using the "one-step deviation principle" of dynamic programming. Note, however, that we need to consider the strategy at every information set of every type of the agent, including the types that are not "legal," i.e., are not supposed to reach this information set given their equilibrium strategies.

[16] Note that we use the interim definition of Bayesian incentive-compatibility: each agent i must be maximizing his expected utility at *all* information sets h, including those at which his type u_i has probability $p^h_i(u_i) = 0$.

the state, agent i will always (weakly) regret not having been truthful. Hence he cannot have a profitable deviation. □

Note that the transfer rule t used in the proposition need not be computed by the protocol — we only need its existence. By the same logic as in Lemma 3, the protocol will then compute some other transfer rule that will satisfy the incentives of the agents at each step. Also, note that to apply Proposition 3, it suffices to check the BIC(p^h) inequalities for each agent i only at his information sets h at which his move may be his last one. This will immediately imply BIC(p^h) at all the other information sets h.

6.2. *An exponential upper bound*

In the same spirit as with ex post implementation, we obtain the following result:

Proposition 4. *Given a BIC(p)-implementable decision rule f for some product distribution p, any simultaneous communication protocol that computes f is BIC(p)-incentivizable.*

Proof. Consider the information partition H of an SBDM with this protocol. With this information partition, since no agent learns anything about the other agents' types during the execution of the protocol, the result follows immediately from Proposition 3 and from the definition of BIC(p)-implementability. □

Corollary 3. *If f is a BIC(p)-implementable decision rule for some product distribution p, then* $\mathrm{CC}_p^{BIC}(f) \leqslant 2^{\mathrm{CC}(f)} - 1$.

Proof. Given the protocol $\mathcal{P}$ that achieves the lower bound $\mathrm{CC}(f)$, by Lemma 1, there is a simultaneous communication protocol $\mathcal{P}'$ that computes the same decision rule such that $d(\mathcal{P}') \leqslant 2^{d(\mathcal{P})} - 1 = 2^{\mathrm{CC}(f)} - 1$. By Proposition 4, $\mathcal{P}'$ is incentivizable, which proves the result. □

This exponential upper bound implies, in particular, that any BIC implementable decision rule f that can be computed with finite communication can also be BIC implemented in a finite BDM (even if the state space U is infinite).[17]

Example 4. Consider the Manager–Expert setting of Example 2. Protocol $\mathcal{P}_0$ has depth $2\lceil \log_2 k \rceil$, so by Corollary 3 there exists a simultaneous communication protocol of depth $2^{2\lceil \log_2 k \rceil} - 1 \sim k^2$ that computes the same decision rule and that is BIC(p)-incentivizable. Note that the communication cost of this protocol is finite whereas full revelation would have required infinite communication (because of the continuous utility range [0, 1] of the Expert).

6.3. *The upper bound is tight*

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

17 Our restriction to independent types and our use of the worst-case communication cost measure are both crucial for this result. With correlated types, or with the average-case communication cost, the BIC overhead can be arbitrarily high, as we show in Appendices B.4, B.1 respectively.

Proposition 5. *For any integer $k > 0$, there exists a BIC(p)-implementable decision rule $f : U \to X$ such that $\mathrm{CC}(f) \leqslant 2\lceil \log_2 k \rceil$ but $\mathrm{CC}_p^{BIC}(f) \geqslant 0.5 \log_2 \binom{k}{k/4}$, with a uniform state distribution p.*

Proof. See Appendix C.1. □

By Stirling's formula, $0.5 \log_2 \binom{k}{k/4}$ is asymptotically equivalent to Ak (where $A \approx 0.41$), which is exponentially higher than $\lceil \log_2 k \rceil$. This shows that our exponential upper bound is essentially tight.

Proposition 5 is proven in Appendix C.1 by formalizing the following example:

Example 5. Consider the Manager–Expert setting of Example 2. The communication complexity of the decision rule is $\mathrm{CC}(f) \leqslant \mathrm{CC}(\mathcal{P}_0) = 2\lceil \log_2 k \rceil$. We have seen in Example 4 that there exists a BIC(p)-incentivizable protocol with a depth of about k^2. We can reduce the communication cost somewhat by using the following BIC(p) BDM: the Expert reports his entire mapping δ, and then the Manager announces the outcome $\delta(m)$. But the communication cost of the Expert reporting his mapping δ is of order $\log_2 k! \sim k \log_2 k$, which is still exponentially higher than $\mathrm{CC}(f)$. We prove that any mechanism satisfying the Expert's Bayesian incentive constraints cannot have a significantly lower communication cost — it must be at least of order k, which is still exponentially higher than $\mathrm{CC}(f)$. The intuition for the proof has two parts: (1) Transfers cannot be used effectively to counterbalance the Expert's private utilities from the outcomes, since the utility range of the Expert is continuous, and so cannot be extracted with finite communication, and (2) Any mechanism with a lower communication cost than of order k must significantly reduce the information revelation by the Expert, which can be achieved only by giving the Expert too much information too soon about the desired consequence m, allowing him to infer how his knowledge will be used and to bias his report according to his preferences.

Proposition 5 demonstrates that the communication cost of selfishness can be prohibitive in the Bayesian setting. After this negative result, we conclude with the following positive one.

6.4. No overhead for EPIC-implementable decision rules

In this case (which includes all efficient decision rules and all "affine maximizers" [4]), there is no communication cost of selfishness.[18]

Proposition 6. *If $\mathcal{P}$ is a protocol that computes an EPIC-implementable decision rule f, then $\mathcal{P}$ is BIC(p)-incentivizable for every product state distribution p.*

Proof. This follows from Proposition 3, as the decision rule f is EPIC-implementable with some particular transfer rule t, and hence f is also BIC(p)-implementable for every distribution p with the same transfer rule t. □

Example 6. Consider the efficient object allocation setting described in Example 1 with some product distribution p over $U = U_1 \times U_2$. Recall that the efficient allocation rule is f EPIC-

[18] While the mechanism constructed in Proposition 6 need not be budget-balanced, it is in fact possible to construct BIC transfers that achieve budget balance (i.e., $\sum_{i \in I} t_i(u) = 0 \ \forall u$) along the lines suggested in [2, Proposition 2].

implementable, e.g., using the Vickrey transfers that charge each agent a price for the object that is equal to the other's reported valuation, and charge nothing when he does not get the object. Recall also that protocol $\mathcal{P}_0$, which computes f without revealing Agent 2's value, is not EPIC-incentivizable. We can still EPIC-incentivize Agent 2 in $\mathcal{P}_0$ by having him pay a price for the object that is equal to Agent 1's announced valuation. As for Agent 1, we cannot EPIC-incentivize him, but we can BIC(p)-incentivize him by having him pay a price for the object that is equal to the *expected* value of Agent 2 conditional on it being below Agent 1's announcement. Intuitively, this price makes him internalize the expected externality he imposes on Agent 2 with his announcement (assuming that Agent 2 reports truthfully).

Proposition 6 also implies that there is no communication cost of selfishness with respect to the *average-case* communication cost measure (defined in Appendix B.1) when we use Bayesian implementation for an EPIC-implementable decision rule.

7. Conclusion

We have examined the communication cost of selfishness in the independent private valuations 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 Bayesian implementation.

Also, we have considered some special cases in which the communication cost of selfishness is low. In the ex post setting, it includes the case where the decision rule is efficient and the utility ranges of the agents have low cardinality. In the Bayesian setting, it includes the case where the decision rule is EPIC-implementable: the communication cost of selfishness in this case is zero. Two other low-overhead cases for the ex post setting are considered in Appendix A.

Finally, in Appendix B we consider some extensions of our basic setting. In particular, we show that the communication cost of selfishness is unbounded when we allow interdependent valuations with ex post implementation or correlated types with Bayesian implementation, or when we consider the average-case communication cost measure with either implementation concept. A similar conclusion for the problem of implementing a decision correspondence rather than decision rule is shown in [8].

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

1. How high is the communication cost of selfishness with ex post 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.

Acknowledgments

The authors gratefully acknowledge the support of the National Science Foundation (grants SES 0318447, ITR 0427770). Preliminary results from this work were reported in an extended abstract [7]. The authors thank Kester Tong for excellent research assistance.

Appendix A. Low-overhead cases for ex post implementation

In this appendix, we consider two cases in which the communication cost of selfishness is low. In the first section, we consider agents whose utilities are linear in a single privately known parameter, and decision rules that are efficient or approximately efficient. In the second section, we consider the case in which there are only two agents whose utilities are given with a fixed precision, and arbitrary EPIC-implementable decision rules.

A.1. Efficient decision rule: single-parameter agents

Here we consider decision rules f that are ϵ-efficient for some $\epsilon \geqslant 0$, i.e., always choose an outcome x that maximizes the sum of utilities $\sum_{i \in I} u_i(x)$ within ϵ. Furthermore, we assume that the agents are "single-parameter," i.e.

Definition 9. Agent i is a single-parameter agent if there exists some set $V_i \subset \mathbb{R}$ and some functions $a_i, b_i : X \to \mathbb{R}$ such that the set of agent i's possible utility functions on X takes the form

$$U_i = \{v_i a_i + b_i\colon\ v_i \in V_i\}.$$

It turns out that we can incentivize single-parameter agents while preserving ϵ-efficiency with an overhead that is at most linear in the communication complexity of the decision rule:

Proposition 7. *Consider an environment with I single-parameter agents. Let f be an ϵ-efficient decision rule. Then there exists an ϵ-efficient decision rule f' for which*

$$\mathrm{CC}^{EPIC}(f') \leqslant I \cdot \big(\mathrm{CC}(f) + 1\big).$$

Proof. Consider a protocol $\mathcal{P}$ computing f. For each node l of the protocol and each agent i, let

$$\underline{v}_i(l) = \inf\big\{v_i \in V_i\colon\ v_i a_i + b_i \in U_i(l)\big\},$$
$$\overline{v}_i(l) = \sup\big\{v_i \in V_i\colon\ v_i a_i + b_i \in U_i(l)\big\}.$$

Given that the agents' utilities are linear in v_i's, the outcome $x(l)$ must be ϵ-efficient for any $(v_1, \ldots, v_I) \in \prod_{i \in I}[\underline{v}_i(l), \overline{v}_i(l)]$. For each agent i, the thresholds $\underline{v}_i(l)$, $\overline{v}_i(l)$ for $l \in L$ partition set V_i into at most $2|L|$ intervals. Consider now the simultaneous communication protocol $\mathcal{P}'$ in which each agent i reports which of these intervals his v_i lies in. This protocol allows us to find a leaf $l \in L$ of $\mathcal{P}$ for which $(v_1, \ldots, v_I) \in \prod_{i \in I}[\underline{v}_i(l), \overline{v}_i(l)]$ and implement the outcome $x(l)$. The new protocol $\mathcal{P}'$ computes an ϵ-efficient outcome, with each agent sending at most $\log_2(2|L|) \leqslant 1 + d(\mathcal{P})$ 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.[19]

A.2. Two agents with fixed utility precision

Recall from (3) that when decision rule f is EPIC-implementable with transfer rule t, 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, we can 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 γ. Then for every EPIC-implementable decision rule f,*

$$\mathrm{CC}^{EPIC}(f) \leqslant \mathrm{CC}(f) + 2(\gamma + 1).$$

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

Appendix B. Extensions

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

B.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 states:

Definition 10. 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 $\mathrm{ACC}_p(\mathcal{P}) = \mathbb{E}_{\tilde{u}}[d(\mathcal{P}, \tilde{u})]$, where $\tilde{u}$ is drawn from p. Furthermore, given a function $f : U \to X$, we define the average communication complexity of f given state distribution p as $\mathrm{ACC}_p(f) = \min_{\mathcal{P}:Fun(\mathcal{P})=f} \mathrm{ACC}_p(\mathcal{P})$.

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

[19] This result could also be derived using a theorem in [3] 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.

The average-case communication cost of selfishness is the difference between $\mathrm{ACC}_p(f)$ and $\mathrm{ACC}_p^{EPIC}(f)$ with ex post implementation, and it is the difference between $\mathrm{ACC}_p(f)$ and $\mathrm{ACC}_p^{BIC}(f)$ with Bayesian implementation.

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

We now show that with ex post implementation, the communication cost of selfishness is unbounded, even if we restrict attention to the case of two agents, an efficient decision rule, and the uniform state distribution.

Proposition 9. *For every $\alpha > 0$ there exists an efficient decision rule f with two agents such that, given the uniform state distribution p: $\mathrm{ACC}_p(f) < 4$ but $\mathrm{ACC}_p^{EPIC}(f) > \alpha$.*

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 $U_1 = U_2 = \{k2^{-\gamma}\colon k = 0, \ldots, 2^\gamma - 1\}$. 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 γ. 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 γ bits. See Appendix C.2 for the complete proof. □

Our average-case analysis can be extended to infinite protocols. While we have not formally defined such protocols, we can imagine that there are 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 average-case communication cost of selfishness is also unbounded for Bayesian implementation, even with only two agents:

Proposition 10. *For any $\alpha > 0$ and $\epsilon > 0$, there exists a BIC(p)-implementable decision rule f with two agents such that: $\mathrm{ACC}_p(f) < 1 + \epsilon$ but $\mathrm{ACC}_p^{BIC}(f) > \alpha$.*

Proof. Consider the rule f used to prove Proposition 5. The rule satisfies $\mathrm{ACC}_p(f) \leqslant \mathrm{CC}(f) \leqslant 2\lceil\log_2 k\rceil$, but also satisfies $\mathrm{ACC}_p^{BIC}(f) \geqslant 0.5 \log_2 \binom{k}{k/4}$, as shown in Appendix C.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 $0.5\epsilon/\lceil\log_2 k\rceil$, and by adding an outcome x_0 that always gives utility 0 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:

$$\mathrm{ACC}_p(f') \leqslant 1 \cdot \bigl(1 - 0.5\epsilon/\lceil\log_2 k\rceil\bigr) + 2\lceil\log_2 k\rceil \cdot 0.5\epsilon/\lceil\log_2 k\rceil < 1 + \epsilon.$$

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

To prove Proposition 10, we constructed an artificial decision rule where, with a high probability, the communication cost is very low. However, Appendix C.1 shows that in a more natural decision rule (described in Example 5) with a uniform probability distribution over types the communication cost of selfishness can be exponential for Bayesian implementation. Also, Proposition 9 above used the standard auction setting with a uniform probability distribution. These cases suggest that the average-case communication cost of selfishness can be severe even in simple and "standard" cases.

B.2. Dominant strategy implementation

Definition 11. 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:

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

Since DIC is stronger than EPIC, 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 a maximally coarse information partition. 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 strategies of an agent to satisfy the incentives *of the agent himself*, whereas in the dominant strategy case, we need to reduce the number of strategies available of an agent to satisfy the incentives *of the other agents*.

B.3. Interdependent values

With interdependent values, an agent's utility function is determined not only by his type, but by the types of the other agents as well. In this case, the overhead may be unbounded for ex post implementation. We illustrate this with the following example.

Example 7. Consider the efficient object allocation setting with interdependent values. For example, the object is a used car that initially belongs to Agent 1. Agent 1's value for this car is his private type $v_1 \in \{k2^{-\gamma}\colon k = 1, \dots, 2^\gamma\}$. Agent 2's type is a bit $c \in \{1, 0\}$ that describes whether he is a mechanic ($c = 1$) or not ($c = 0$). If $c = 1$, Agent 2 is able to repair the car, and his value for it is $v_2^1 = v_1 + 2^{-\gamma}$. However, if $c = 0$, Agent 2 cannot repair the car, and his valuation for it is $v_2^0 = v_1 - 2^{-\gamma}$. The efficient outcome is to give the car to Agent 2 if and only if he is a mechanic ($c = 1$), and with honest agents, this outcome can be computed with a fixed communication of just 1 bit: Agent 2 reports his type c. If agents are selfish, the rule is still EPIC-implementable with full revelation and the following transfer rule: Agent 1 never pays or receives any transfer, but Agent 2 must pay v_1 (to a third party) if he gets the car (i.e., if he announces that he is a mechanic). However, any protocol that is EPIC-incentivizable needs to reveal Agent 1's value v_1 within $2^{-\gamma}$ to satisfy the ex post incentive constraints of Agent 2, which takes at least $\gamma - 1$ bits. (This can be made formal along the lines of the proof of Proposition 9.) Hence the communication cost of selfishness is arbitrarily high in this case.

As far as *Bayesian* implementation is concerned, however, we still have an exponential upper bound on the overhead, provided that the types are independently distributed. Indeed, Propositions 3, 4 and Corollary 3 all hold for this case.

B.4. 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 state distribution. As for Bayesian implementation, the communication cost of selfishness with correlated types may be unbounded:

Example 8. There are two agents, and we consider the incentives of only Agent 1, who privately knows a string $w \in (0,1)^m$. The desired outcome is the parity of the string, i.e., $f(w) = (\sum_{j=1}^{m} w(j)) \bmod 2$. Agent 1 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 k between 1 and m, and the value of $w(k)$. The distribution of k 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 $\mathrm{BIC}(p)$ with a high monetary punishment for Agent 1 "caught" lying, i.e., announcing a wrong value of $w(k)$. However, any $\mathrm{BIC}(p)$-incentivizable protocol must have depth at least $\log_2 m$, as otherwise Agent 1 would have fewer than 2^m strategies, and hence there would be two 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 to prevent Agent 1, when he has type w'' and prefers outcome 0, from choosing the strategy of types w and w' (without preventing w or w' from being truthful). Hence the communication cost of selfishness can be made arbitrarily high by choosing m large enough.

We can attribute this increase in the overhead to the failure of Lemma 3 with correlated types: we cannot stop a $\mathrm{BIC}(p)$-incentivizable protocol when the outcome is computed, and hence the computation of the transfers may cause an increase in the communication requirements. Intuitively, this example offers one reason why surplus-extraction mechanisms for correlated types proposed by Cremer and McLean [6] may not be practical: in some cases, such mechanisms may require prohibitive communication.

Appendix C. Technical proofs

We begin with a simple lemma that is useful for bounding below the average-case communication complexity $\mathrm{ACC}_p(\mathcal{P})$ of a protocol $\mathcal{P}$ (as defined in Appendix B.1):

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

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 ([25], surveyed in [5]) implies that $\mathrm{ACC}_p(\mathcal{P})$ is bounded below by the entropy of $\tilde{l}$, defined as $\mathrm{H}(\tilde{l}) = -\sum_{l\in L} p(l)\log_2 p(l)$.

Under our assumptions, this entropy is in turn bounded below as follows:

$$\mathrm{H}(\tilde{l}) \geqslant -\sum_{l\in L'} p(l)\log_2 p(l) \geqslant -\sum_{l\in L'} p(l)\log_2 \beta \geqslant -\alpha\log_2\beta. \qquad \square$$

C.1. Proof of Proposition 5 (exponential overhead for Bayesian implementation)

Consider the Manager–Expert setting of Examples 2 and 5. Recall from Example 5 that $\mathrm{CC}(f) \leqslant 2\lceil \log_2 k\rceil$, and that f is BIC(p)-implementable. We now prove that $\mathrm{CC}_p^{BIC}(f) \geqslant 0.5\log_2\binom{k}{k/4}$ by proving the stronger statement $\mathrm{ACC}_p^{BIC}(f) \geqslant 0.5\log_2\binom{k}{k/4}$. (Where $\mathrm{ACC}_p^{BIC}(f)$ is the average Bayesian incentive communication complexity, as defined in Appendix B.1.)

First we observe that in a finite BDM, the Expert can communicate "little information" about u. Formally, even if we execute the BDM for all possible values of δ and m, at the conclusion the set $U' \subset U$ of still possible values of u must have a positive measure with probability 1. For any such positive-measure set U', we can restrict attention to the protocol's execution when we know that $u \in U'$ (in which case the protocol reveals no further information about u). Informally, this corresponds to a modified communication protocol in which the Expert announces upfront that $u \in U'$ (and this announcement is not counted towards the communication cost), and then the agents proceed to communicate information about δ and m. We will bound below the protocol's average communication complexity conditional on any positive-measure set U', and this will also bound unconditional average communication complexity.

Now, let us focus on the set H_1 of the Expert's infosets. For each infoset $h \in H_1$, let $M_h \subset M$ denote the set of legal m's at h — i.e., those the Expert still considers possible at h. (Thus, M_h is the union of the legal m's at all nodes in h.) Let Δ_h denote the set of legal δ's at infoset h — i.e., those for which the Expert can arrive at h (which must be the same in every node in h, since the Expert cannot distinguish among them).[20] Since information about m is communicated by the Manager and information about Δ is communicated by the Expert, the set of legal (m,δ) pairs at infoset h must be $M_h \times \Delta_h$.

We will say that "at infoset $h \in H_1$, the Expert has revealed the image of set $\overline{M} \subset M$" if $\delta(\overline{M})$ has the same value for all $\delta \in \Delta_h$.

Claim 1. *If $\mathcal{B}$ is a BIC(p) BDM implementing f, then at each Expert's information set $h \in H_1$ reached with a positive probability he has revealed the image of M_h.*

Proof. Consider any infoset $h \in H_1$ at which the Expert has not revealed the image of M_h, i.e., there exist $\delta', \delta'' \in \Delta_h$ such that $\delta'(M_h) \neq \delta''(M_h)$. Starting from h, by reporting according to δ' the Expert would induce a uniform probability distribution over outcomes from $\delta'(M_h)$, while by reporting according to δ'' he would induce a uniform probability distribution over outcomes

[20] It is important to keep in mind that in our terminology the protocol does not include the "moves of nature" informing agents about their types, and therefore the Expert's information sets in the protocol only describe information revealed through the agents' moves and do not reveal the Expert's observation of his own type.

from $\delta''(M_h)$. For the Expert to report truthfully regardless of whether his true type is (δ', u) or (δ'', u), the difference between his expected transfers in the two cases must equal

$$F(u) \equiv \frac{1}{|M_h|}\left[\sum_{x\in\delta'(M_h)} u(x) - \sum_{x\in\delta''(M_h)} u(x)\right] = \frac{1}{|M_h|}\left[\sum_{x\in\delta'(M_h)\setminus\delta''(M_h)} u(x) - \sum_{x\in\delta''(M_h)\setminus\delta'(M_h)} u(x)\right].$$

Since the difference between the two expected transfers must be known at the infoset, and the expert should report truthfully for all $u \in U'$, $F(u)$ must be constant on $u \in U'$. But since $\delta'(M_h) \neq \delta''(M_h)$, the sums in the last expression contain at least one term and so any set on which $F(u)$ is constant must be a zero-measure set. □

Claim 2. *In any BDM that implements f in BIC(p), the Expert must with probability* $1/2$ *eventually reveal the image of a set of size between $k/4$ and $k/2$.*

Proof. We show that the probability is at least $1/2$ conditional on any fixed $\delta = \hat{\delta}$, which will imply that the unconditional probability is at least $1/2$ as well. Construct a tree $T(\hat{\delta})$ (not necessarily binary) consisting of the Expert's infosets from H_1 that he could possibly visit when $\delta = \hat{\delta}$ and he reports truthfully (i.e., those infosets $h \in H_1$ at which $\hat{\delta} \in \Delta_h$). For any given $h \in T(\hat{\delta})$, let *child*$(h) \subset T(\hat{\delta})$ denote the set of children of h in $T(\hat{\delta})$ (i.e., the set of infosets in $T(\hat{\delta})$ that can be visited by the Expert immediately after visiting h). (That this child relation must induce a tree follows from the Expert's perfect recall.)

Let us walk from the root of $T(\hat{\delta})$ down a path in the tree while always choosing the child h that has the largest $|M_h|$. We continue until we get to a first node h all of whose children h' have $|M_{h'}| < k/2$. By construction we must have $|M_h| \geqslant k/2$, and so the protocol's execution will go through h with probability $|M_h|/|M| \geqslant 1/2$ when $\delta = \hat{\delta}$. Now, we should be able to select a subset $H' \subset$ *child*(h) such that the size of the set $\overline{M} \equiv \bigcup_{h'\in H'} M_{h'}$ is between $k/4$ and $k/2$. (Indeed, if there exists $h' \in$ *child*(H) with $|M_h| \geqslant k/4$ then we can take $H' = \{h'\}$, otherwise we can keep adding elements of *child*(H) to H' until $|\overline{M}|$ first exceeds $k/4$, in which case it still falls short of $k/2$.)

Recall from Claim 1 that at any $h' \in$ *child*(h) the Expert must have revealed the image of $M_{h'}$. This revelation must have happened with the Expert's move at h. Formally, all infosets $h' \in$ *child*(h) have the same history of the Expert's moves (the same moves that the Expert of type $\hat{\delta}$ would make), hence they all have the same $\Delta_{h'}$, and so Claim 1 implies that at *each* of them the Expert has revealed the image of $M_{h''}$ for *any* $h'' \in$ *child*(h), and therefore the image of $\overline{M}$ as well. Thus, the image of $\overline{M}$ is revealed whenever the execution of the protocol passes through infoset h, and we have already established that this occurs with probability at least $1/2$ when $\delta = \hat{\delta}$. □

Claim 3. *Any protocol in which with probability* $1/2$ *the Expert reveals the image of a set of size between $k/4$ and $k/2$ must have an average communication cost of at least* $0.5\log_2\binom{k}{k/4}$.

Proof. If at an infoset $h \in H_1$ the Expert has revealed the image of set $\overline{M} \subset M$, then $|\Delta_h| \leqslant |\overline{M}|!|M\setminus\overline{M}|!$ If we know that $k/4 \leqslant |\overline{M}| \leqslant k/2$, then $|\Delta_h| \leqslant (3k/4)!(k/4)!$. Since all the δ's are equiprobable, the probability of the protocol's passing through h is at most $|\Delta_h|/k! \leqslant \binom{k}{k/4}^{-1}$, which also bounds above the probability of arriving at any given leaf following h. Thus, with probability $1/2$ the protocol must arrive at a leaf whose probability is at most $\binom{k}{k/4}^{-1}$, hence by Lemma 4, the average communication cost of the protocol is at least $0.5\log_2\binom{k}{k/4}$. □

Thus, we have established a lower bound of $0.5\log_2\binom{k}{k/4}$ on the average communication complexity of any BDM that BIC implements f. (Actually, this is a lower bound on the expected number of bits send by the Expert regarding δ; the proof does not count the information that the Expert might give about his valuations u or the Manager's messages about m.) So $\mathrm{ACC}_p^{BIC}(f) \geqslant 0.5\log_2\binom{k}{k/4}$, which also implies that $\mathrm{CC}_p^{BIC}(f) \geqslant 0.5\log_2\binom{k}{k/4}$.

C.2. Proof of Proposition 9 (unbounded average-case overhead for ex post implementation)

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 $U_1 = U_2 = U = \{k2^{-\gamma}\colon k = 0, \ldots, 2^\gamma - 1\}$. 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 [1] and [13]: At each round $m = 1, \ldots, \gamma$, each agent i reports the mth bit in the binary expansion of his valuation u_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 γ 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 $1/2$. Therefore, the expected number of rounds is at most 2, and so the average-case communication complexity is at most 4, for any γ.

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 $\tau_2(f(u_1, u_2), u_1)$. Furthermore, EPIC inequalities imply that

$$\left|\tau_2(2, u_1) - \tau_2(1, u_1) - u_1\right| \leqslant 2^{-\gamma} \quad \text{for every } u_1 \in \left(0, 1 - 2^{-\gamma}\right), \tag{5}$$

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 $\gamma \geqslant 3$. 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 $u_2 \geqslant 3/4$, Agent 3 has type $u_3 < 1/4$, and Agent 1 has type $u_1 \in [1/4, 3/4)$ (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 (5) the difference between Agent 2's and Agent 3's transfers pins down the realization of u_1 within $2^{-\gamma}$. Thus, in this event, each outcome 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 γ to get an arbitrarily high average-case communication cost.

References

[1] K.J. Arrow, L. Pesotchinsky, M. Sobel, On partitioning a sample with binary-type questions in lieu of collecting observations, J. Amer. Statistical Assoc. 76 (1981) 402–409.

[2] S. Athey, I. Segal, An efficient dynamic mechanism, Working paper, 2006.

[3] L. Blumrosen, N. Nisan, I. Segal, Auctions with severely bounded communication, J. Artif. Intell. Res. 28 (2007) 233–266.

[4] S. Bikhchandani, S. Chatterji, R. Lavi, A. Mu'alem, N. Nisan, A. Sen, Weak monotonicity characterizes deterministic dominant strategy implementation, Econometrica 74 (4) (2006) 1109–1132.

[5] T.M. Cover, J.A. Thomas, Elements of Information Theory, John Wiley & Sons, Inc., New York, 1991.

[6] J. Cremer, R.P. McLean, Full extraction of the surplus in Bayesian and dominant strategy auctions, Econometrica 56 (1988) 1247–1257.

[7] R. Fadel, I. Segal, The communication cost of selfishness: Ex post implementation, in: Proceedings of the 10th Conference in Theoretical Aspects of Rationality and Knowledge, 2005.

[8] R. Fadel, A contribution to canonical hardness, in: The 7th International Workshop on Agent-Mediated Electronic Commerce (AMEC), 2005.

[9] J. Feigenbaum, A. Krishnamurthy, R. Sami, S. Shenker, Hardness results for multicast cost sharing, Theoret. Comput. Sci. 304 (2003) 215–236.

[10] F. Forges, Equilibria with communication in a job market example, Quart. J. Econ. 105 (1990) 375–398.

[11] D. Fudenberg, J. Tirole, Game Theory, MIT Press, Cambridge, MA, 1991.

[12] J. Green, J.-J. Laffont, Limited communication and incentive compatibility, in: T. Groves, R. Radner, S. Reiter (Eds.), Information, Incentives, and Economic Mechanisms: Essays in Honor of Leonid Hurwicz, University of Minnesota Press, 1987.

[13] E. Grigorieva, P.J.-J. Herings, R. Muller, D. Vermeulen, The private value single item bisection auction, Econ. Theory 30 (2007) 107–118.

[14] R. Johari, Efficiency loss in market mechanisms for resource allocation, PhD thesis, Massachusetts Institute of Technology, 2004.

[15] E. Kushilevitz, N. Nisan, Communication Complexity, Cambridge University Press, 1997.

[16] S. Lahaie, D. Parkes, Applying learning algorithms to preference elicitation, in: Proceedings of the 5th ACM conference on Electronic Commerce, 2004.

[17] A. Mas-Colell, M. Whinston, J.R. Green, Microeconomic Theory, Oxford University Press, New York, 1995.

[18] N. Melumad, D. Mookherjee, S. Reichelstein, A theory of responsibility centers, J. Acc. Econ. 15 (1992) 445–484.

[19] D. Mookherjee, Decentralization, hierarchies and incentives: A mechanism design perspective, J. Econ. Lit. 44 (2006) 367–390.

[20] R.B. Myerson, Game Theory: Analysis of Conflict, Harvard University Press, Cambridge, MA, 1991.

[21] N. Nisan, I. Segal, The communication requirements of efficient allocations and supporting prices, J. Econ. Theory 129 (2006) 192–224.

[22] S. Reichelstein, Incentive compatibility and informational requirements, J. Econ. Theory 34 (1984) 32–51.

[23] K. Roberts, The characterization of implementable choice rules, in: J.-J. Laffont (Ed.), Aggregation and Revelation of Preferences, North-Holland, 1979.

[24] I. Segal, The communication requirements of social choice rules and supporting budget sets, J. Econ. Theory 136 (2007) 341–378.

[25] C.E. Shannon, A mathematical theory of communication, Bell System Tech. J. 27 (1948) 379–423, 623–656.

[26] A.C.-C. Yao, Some complexity questions related to distributive computing (preliminary report), in: Proceedings of the 11th Annual ACM Symposium on Theory of Computing, ACM Press, 1979, pp. 209–213.