

NOTES AND COMMENTS

OVERCOMING INCENTIVE CONSTRAINTS BY  
LINKING DECISIONS<sup>1</sup>

BY MATTHEW O. JACKSON AND HUGO F. SONNENSCHIN<sup>2</sup>

Consider a Bayesian collective decision problem in which the preferences of agents are private information. We provide a general demonstration that the utility costs associated with incentive constraints become negligible when the decision problem is linked with a large number of independent copies of itself. This is established by defining a mechanism in which agents must budget their representations of preferences so that the frequency of preferences across problems mirrors the underlying distribution of preferences, and then arguing that agents' incentives are to satisfy their budget by being as truthful as possible. We also show that all equilibria of the linking mechanisms converge to the target utility levels. The mechanisms do not require transferable utility or interpersonal comparisons of utility, and are immune to manipulations by coalitions.

KEYWORDS: Mechanism design, implementation, linking, Bayesian equilibrium, efficiency.

1. INTRODUCTION

OVER THE PAST FIFTY YEARS we have learned that social welfare possibilities depend not only on resources and technology, but also on incentive constraints (including participation constraints) and the ability of social institutions to mediate those constraints. Thus, voting systems, labor contracts, financial contracts, auction forms, and a host of other practical arrangements are now commonly formulated as Bayesian games and are judged in terms of their ability to mediate incentive constraints.

This paper demonstrates how the limitations that incentive constraints impose on the attainment of socially efficient outcomes disappear when problems are linked. We exploit the idea that when independent social decision problems are linked, it makes sense to speak of “rationing” or “budgeting” an agent's representations. In more formal language, we consider an abstract Bayesian collective decision problem and an *ex ante* Pareto efficient social choice function  $f$  that indicates the collective decision we would like to make as a function

<sup>1</sup>This paper supercedes “The Linking of Collective Decisions and Efficiency,” Caltech Social Science Working Paper 1159, March 2003. A longer version of this paper with additional examples, discussion, and results can be found at <http://www.stanford.edu/~jacksonm/linking.pdf>.

<sup>2</sup>Financial support under NSF Grants SES-9986190 and SES-0316493 is gratefully acknowledged, as is support from the University of Chicago, where Jackson was visiting when this research began. We thank Zachary Cohn and Phil Reny for calling our attention to the treatment of participation constraints, Bill Zame for questions about coalitional considerations, and Robert Veszteg and Hitoshi Matsushima as discussants of the paper. We also thank the editor, three anonymous referees, Mahmut Erdem, Drew Fudenberg, Eric Maskin, and Laurent Mathevet for their comments and suggestions.

of the realized preferences of the  $n$  agents. Let  $(u_1, u_2, \dots, u_n)$  denote the ex ante expected utilities that are achieved under  $f$ . This  $f$  will generally not be implementable because of incentive constraints. Now, consider  $K$  copies of the decision problem, where agents' preferences are additively separable and independently distributed across the problems. We show that as  $K$  becomes large it is essentially possible to implement  $f$  on each problem and thus achieve the target utilities  $(u_1, u_2, \dots, u_n)$  on each problem.

We establish this result by constructing a mechanism in which each agent announces a  $K$ -vector of preferences. The announcements of the agents are "budgeted" so that the distribution of types across problems must mirror the underlying distribution of their preferences. The decision on each of the  $K$  problems is made according to the desired  $f$  as if the announcements were true. With some adjustments in this idea to deal with the case of more than two agents (as well as participation constraints), we show that in the limit there is no gain from lying. In fact, for *every*  $K$  there is an equilibrium in which all agents are telling the truth as fully as the constraint on their announcements permits. Moreover, we show that being as truthful as possible secures for any agent an expected utility that for large  $K$  approximates the target utility level *regardless of the strategies followed by the other agents*. Thus, we can conclude that *all* equilibria of the linking mechanisms converge to the target utility levels. Furthermore, our mechanisms do not require any transferable utility or interpersonal comparisons of utility, and they are also immune to manipulations by coalitions. Thus, the machinery that is generally used in mechanism design theory and relies heavily on transferable utility is not used here.

The closest antecedents of our work are Townsend (1982) and McAfee (1992).<sup>3,4</sup> Townsend (1982) examined a repeated risk-sharing problem between a risk-neutral agent with a constant endowment and a risk-averse agent with a risky endowment that takes on either a high or low value. He noted that by limiting the number of times the risk-averse agent can claim to have a

<sup>3</sup>Radner (1981) examined repeated contracting in a finitely repeated principal-agent setting, but relies on  $\varepsilon$ -equilibrium and trigger strategies that are closer in spirit to the incomplete information repeated games literature (e.g., Green and Porter (1984), Abreu, Pearce, and Stacchetti (1991)) than the approach taken here. Equilibria in many incomplete information repeated games are bounded away from efficiency (e.g., see Fudenberg and Levine (1994)).

<sup>4</sup>A short list of other research that involves the linking of decisions includes logrolling (e.g., Tullock (1970), Wilson (1969), Miller (1977)), bundling of goods by a monopolist (Adams and Yellen (1976), McAfee, McMillan, and Whinston (1979), Chakraborty and Harbaugh (2003), Armstrong (1999), Bakos and Brynjolfsson (1999, 2000)), agency problems (Maskin and Tirole (1990)), and subsequent papers by Fang and Norman (2003), who examined the efficiency gains from the bundling of excludable public goods, and Hörner and Jamison (2004), who examined collusion in a repeated Bertrand game. Although some of these papers also rely on laws of large numbers (e.g., Armstrong (1999), Bakos and Brynjolfsson (1999, 2000)), our approach differs in showing that ex ante Pareto efficiency of a social choice function or allocation rule can be used to give agents incentives to report their types as truthfully as possible and, moreover, that this applies to any collective decision problem.

low endowment, an approximately efficient outcome can be reached. McAfee (1992) examined a group of agents who allocate a set of indivisible objects when agents may have different valuations for different objects and the objective is to get objects to the agents who value them most. He showed, under some symmetry assumptions, that a mechanism where agents take turns selecting objects approaches full efficiency as the number of objects grows.<sup>5</sup> Our mechanism simplifies to Townsend's in his special context. McAfee's mechanism, although different in structure, would lead to approximately the same outcomes as a version of our linking mechanism that sought to give objects to agents with the highest valuation. However, their results give little indication of the shape of the general theory presented here, especially when no transfers are present.<sup>6</sup>

Our results show that if linking is possible, then efficiency is generally obtainable, even without any transferability, and by a simple mechanism with a number of desirable properties. These results can be thought of as stemming from two insights. The first, which appears in many forms in the previous literature, is that when many problems are linked, laws of large numbers tell us that the realized frequency of players' types will closely match the underlying distribution. The second insight is the more innovative aspect of our results and the key to their general coverage. It is that if the desired social choice function is *ex ante* Pareto efficient and we employ a simple mechanism that budgets agents' announcements to match the underlying distribution, then agents have an incentive to be as truthful as possible.

## 2. EXAMPLES

**EXAMPLE 1 —A Voting Problem:** Consider a two-agent society that makes a binary decision represented by  $d \in \{a, b\}$ . An agent's preferences are summarized by the difference in utilities between decisions  $a$  and  $b$ , denoted  $v_i = v_i(b) - v_i(a)$ . For simplicity, assume that each  $v_i$  is independently and identically distributed and takes on values in  $\{-2, -1, 1, 2\}$  with equal probability.

Suppose that we wish to choose the decision that maximizes the sum of the utilities and in the case of a tie, we flip a coin. Clearly this social choice function is not implementable. The unique incentive compatible social choice function

<sup>5</sup>A related mechanism was discussed by Pesendorfer (2000) in the context of a set of bidders who collude in a sequence of auctions trying to decide who should win which auctions. See also Blume and Heidhues (2002), Campbell (1998), and Chakraborty, Gupta, and Harbaugh (2002), as well as earlier work on multimarket collusion by Bernheim and Whinston (1990).

<sup>6</sup>Our own investigations were spurred by trying to understand whether full efficiency could be achieved in a simple voting setting, after seeing the creative and innovative storable votes mechanism of Casella (2002) (see also Hortala-Vallve (2003) and Casella and Palfrey (2003)). Casella showed that in some cases, an equilibrium of the storable votes mechanism offers a Pareto improvement over separate votes.

that is anonymous and neutral, and maximizes total utility subject to incentive constraints, corresponds to having agents vote over the alternatives and flipping a coin in the event of a tie (a version of May's (1952) theorem). The inefficiency is that we are not able to discover agents' intensity of preference in the event of a tied vote. This is not an issue of *interpersonal* comparisons, but rather *intrapersonal* comparisons. An agent would be better off if he won the ties when he was of type  $-2$  or  $2$ , at the cost of losing the ties when of type  $-1$  or  $1$ . However, this cannot be achieved in an incentive compatible way with just one decision, because the agent would always pretend to be of the high type. However, if two such decisions are linked, we could, for instance, ask the agents to declare that they are of a high type on just one of the two decisions. Essentially, by linking the decisions together, we can ask, "Which decision do you care more about?" This can be answered in an incentive compatible way in the linked problem. Effectively, linking the problem changes *ex ante* inefficiencies—"I would like to make trades over my different possible future selves,"—to *ex post* inefficiencies—"I now actually have different selves and would be happy to make trades across them."

If we link a number  $K$  (say divisible by 4) of independent decisions together of the foregoing type, then we can budget each agent to announce  $-2$  on  $K/4$  problems,  $-1$  on  $K/4$  problems, etc. Now let us simply choose outcomes according to the desired social choice function, treating the agents' announcements as if they were truthful. It turns out that it is in the agents' interest to be as truthful as they can. Clearly, if an agent has a positive type on problem  $k$  and a negative type on problem  $k'$ , then the agent would not gain by switching those announcements. However, if the agent has type 1 on problem  $k$  and type 2 on problem  $k'$ , then does the agent have any incentive to lie and reverse the types in his announcements? If the agent does not have substantially different beliefs about what the other agent's type announcement is likely to be across these problems, then he or she would rather have the choice made in his or her favor on the problem where he or she really is a 2; thus, announcing these types correctly is better than reversing them. There is more work to be done to prove that being as truthful as possible is a good strategy (and essentially the only equilibrium strategy) for this problem, but the essence is clear. With large  $K$ , the mechanism will force the agents to "lie" on a negligible fraction of problems. So, on each problem, we are converging to truthful revelation and the associated *ex ante* (and thus interim and *ex post*) efficient decisions.<sup>7</sup>

Note that this mechanism works without any transfers. If one could use transfers, then there are other ways to mitigate the inefficiency problem in this particular voting problem. However, allowing for transfers cannot always reconcile incentive constraints with efficiency. We now present another example, where even in the presence of transfers there is a conflict between efficiency, incentive compatibility, and participation constraints.

<sup>7</sup>One can get a faster rate of convergence with modifications to the mechanism.

EXAMPLE 2 —A Bargaining Problem: There is a seller of an object with valuation chosen uniformly from  $\{0.1, 0.3, 0.5, 0.7, 0.9\}$  and a buyer with valuation chosen uniformly from  $\{0.2, 0.4, 0.6, 0.8, 1\}$ . Each agent's utility is the value of the object if he or she has it at the end of the period, net of any transfers that are made. The values are independent and we wish to have the agents trade the object if and only if the buyer's value exceeds that of the seller. To set a target, consider the social choice function that trades the object precisely in these cases and at a price that is the average of the valuations. This mechanism fails to be incentive compatible. In fact, there is no incentive compatible, Pareto efficient, and individually rational mechanism for this problem, as we know from Myerson and Satterthwaite (1983).

As before, we link  $K$  (say divisible by 5) decision problems by requiring each agent to specify exactly  $1/5$  of the problems where they have each valuation. Then we determine the outcomes by using the target outcome function on each problem. As before, there is an approximately truthful equilibrium where agents tell the truth to the maximal extent possible, given that it is possible that they will not have a given valuation on exactly  $1/5$  of the problems. Again, for large  $K$ , the fraction of problems where the correct decision is made goes to 1 in probability.

### 3. A GENERAL THEOREM ON LINKING DECISIONS

#### *Decision Problems*

An  $n$ -agent decision problem is a triple  $\mathcal{D} = (D, U, P)$ . Here  $D$  is a finite set of possible alternative decisions;  $U = U_1 \times \cdots \times U_n$  is a finite set of possible profiles of utility functions  $(u_1, \dots, u_n)$ , where  $u_i: D \rightarrow \mathbb{R}$ ; and  $P = (P_1, \dots, P_n)$  is a profile of probability distributions, where  $P_i$  is a distribution over  $U_i$ .<sup>8</sup> We abuse notation and write  $P(u)$  for the probability of  $u$ , and we assume that the  $u_i$ 's are drawn independently.<sup>9</sup>

#### *Social Choice Functions*

A *social choice function* on a social decision problem  $\mathcal{D} = (D, U, P)$  is a function  $f: U \rightarrow \Delta(D)$ , where  $\Delta(D)$  denotes the set of probability distributions on  $D$ . This is interpreted as the target outcome function. We allow  $f$ 's

<sup>8</sup>Our setting is thus one of private values. If values are interdependent (but still independently distributed), then some of our results still go through whereas others do not. With interdependent valuations there still exists an approximately truthful equilibrium to our mechanism (which extends as is), but for some settings there may also exist other equilibria that are far from being approximately truthful both in announcements and outcomes.

<sup>9</sup>Although restrictive, this requirement ensures that our efficiency results are not obtained by learning something about one agent's type through the reports of others (for instance, as in Crémer and McLean (1988)).

to randomize over decisions because such randomizations admit tie-breaking rules that are common and sometimes needed to define attractive (e.g., neutral and/or anonymous) social choice functions. Let  $f_d(u)$  denote the probability of choosing  $d \in D$ , given the profile of utility functions  $u \in U$ .

A social choice function  $f$  on a decision problem  $\mathcal{D} = (D, U, P)$  is *ex ante Pareto efficient* if there does not exist any social choice function  $f'$  on  $\mathcal{D} = (D, U, P)$  such that

$$\sum_u \left[ P(u) \sum_d (f'_d(u) u_i(d)) \right] \geq \sum_u \left[ P(u) \sum_d (f_d(u) u_i(d)) \right]$$

for all  $i$  with strict inequality for some  $i$ .

### Linking Mechanisms

Given a base decision problem  $\mathcal{D} = (D, U, P)$  and a number  $K$  of linkings, a *linking mechanism*  $(M, g)$  is a message space  $M = M_1 \times \dots \times M_n$  and an outcome function  $g: M \rightarrow \Delta(D^K)$ . In the linking mechanisms that we use to prove our results,  $M_i$  consists of announcements of utility functions for each decision problem. Let  $g_k(m)$  denote the marginal distribution under  $g$  on the  $k$ th decision, where  $m \in M$  is the profile of messages selected by the agents.

When we link  $K$  versions of a decision problem  $\mathcal{D} = (D, U, P)$ , an agent's utility over a set of decisions is simply the sum of utilities. So, the utility that agent  $i$  gets from decisions  $(d^1, \dots, d^K) \in D^K$  given preferences  $(u_i^1, \dots, u_i^K) \in U_i^K$  is given by  $\sum_k u_i^k(d^k)$ . We assume that the randomness is independent across decision problems. Given independence and additive separability, there are absolutely no complementarities across the decision problems, and so any improvements in efficiency obtained through linking must come from being able to trade decisions off against each other to uncover intensities of preferences.

A *strategy* for agent  $i$  in a linking mechanism  $(M, g)$  on  $K$  copies of a decision problem  $\mathcal{D} = (D, U, P)$  is a mapping  $\sigma_i^K: U_i^K \rightarrow \Delta(M_i)$ . We consider Bayesian equilibria of such mechanisms. Given a decision problem  $\mathcal{D} = (D, U, P)$  and a social choice function  $f$  defined on  $\mathcal{D}$ , we say that a sequence of linking mechanisms defined on increasing numbers of linked problems,  $\{(M^1, g^1), (M^2, g^2), \dots, (M^K, g^K), \dots\}$ , and a corresponding sequence of Bayesian equilibria,  $\{\sigma^K\}$ , *approximate*  $f$  if

$$\lim_K \left[ \max_{k \leq K} \text{Prob} \{ g_k^K(\sigma^K(u)) \neq f(u^k) \} \right] = 0.$$

### Strategies that Secure a Utility Level

Given a mechanism  $(M, g)$  on  $K$  linked decision problems, a strategy  $\sigma_i: U_i^K \rightarrow M_i$  *secures a utility level*  $\bar{u}_i$  for agent  $i$  if, for all strategies of the other

agents  $\sigma_{-i}$ ,

$$E \left[ \sum_{k \leq K} u_i(g^k(\sigma_i, \sigma_{-i})) \right] \geq K\bar{u}_i.$$

*The General Linking Mechanisms: A First Look*

Consider  $K$  linked problems. Each agent announces utility functions for the  $K$  problems, as in a direct revelation mechanism. However, the agent’s announcements across the  $K$  problems must match the expected frequency distribution. That is, the number of times that  $i$  can (and must) announce a given utility function  $u_i$  is  $K \times P_i(u_i)$ . With a finite set of problems,  $K \times P_i(u_i)$  may not be integer valued for all  $i$  and  $u_i$ , and so we approximate  $P_i$ . The choice is then made according to  $f$  based on the announcements.

More formally, our  $K$ th linking mechanism,  $(M^K, g^K)$ , is defined as follows: Find any approximation  $P_i^K$  to  $P_i$  such that  $P_i^K(v_i)$  is a multiple of  $\frac{1}{K}$  for each  $v_i \in U_i$ , and the Euclidean distance between  $P_i^K$  and  $P_i$  (viewed as vectors) is minimized.

Agent  $i$ ’s strategy set is

$$M_i^K = \{ \hat{u}_i \in U_i^K \text{ s.t. } \#\{k : \hat{u}_i^k = v_i\} = P_i^K(v_i)K \text{ for each } v_i \in U_i \}.$$

The decision of  $g^k$  for the problem  $k$  is simply  $g^k(m) = f(\hat{u}^k)$ , where  $\hat{u}_i^k$  is  $i$ ’s announced utility function for problem  $k$  under the realized announcement  $m = \hat{u}$ . (The mechanism here is refined in the proofs, because modifications are needed to rule out collusive equilibria in the case of more than three agents.)

The constraint of announcing a distribution of utility functions that approximates the true underlying distribution of types will sometimes force an agent to lie about his utility functions on some problems, because his realizations of utility functions across problems may not have a frequency that is precisely  $P_i$ . Nevertheless, strategies that are as truthful as possible subject to the constraints turn out to be useful strategies for the agents to employ, and so we give such strategies a name. An agent follows a strategy that is *approximately truthful* if the agent’s announcements always involve as few lies as possible. Formally,  $\sigma_i^K : U_i^K \rightarrow M_i^K$  is *approximately truthful* if

$$\#\{k \mid [\sigma_i^K(u_i^1, \dots, u_i^K)]^k \neq u_i^k\} \leq \#\{k \mid m_i^k \neq u_i^k\}$$

for all  $m_i \in M_i^K$  and all  $(u_i^1, \dots, u_i^K) \in U_i^K$ .

*A Theorem on Approximating Efficient Decisions Through Linking*

Let  $\bar{u}_i = E[u_i(f(u))]$  and let  $\bar{u} = (\bar{u}_1, \dots, \bar{u}_n)$  denote the ex ante expected utility levels under the target social choice function. These are the targets for the utility level that we would like to implement.

**THEOREM 1:** *Consider a decision problem  $\mathcal{D}$  and an ex ante Pareto efficient social choice function  $f$  defined on it. There exists a sequence of linking mechanisms  $(M^K, g^K)$  on linked versions of the decision problem such that the following statements hold:*

(i) *There exists a corresponding sequence of Bayesian equilibria that are approximately truthful.*

(ii) *The sequence of linking mechanisms together with these corresponding equilibria approximate  $f$ .*

(iii) *Any sequence of approximately truthful strategies for an agent  $i$  secures a sequence of utility levels that converge to the ex ante target level  $\bar{u}_i$ .*

(iv) *All sequences of Bayesian equilibria of the linking mechanisms result in expected utilities that converge to the ex ante efficient profile of target utilities of  $\bar{u}$  per problem.*

(v) *For any sequence of Bayesian equilibria and any sequence of deviating coalitions, the maximal gain by any agent in the deviating coalitions vanishes along the sequence.*

We remark that the ex ante Pareto efficiency of the social choice function is essential to the result. There are two aspects to the proof. One is that with large numbers of linked problems, there is a high probability that the realized distribution of agents' types will closely match the underlying distribution. This is a standard result of the law of large numbers that has been exploited in the literature in many ways before. The second aspect is establishing that agents have an incentive to be as truthful as possible when faced with our mechanism. This is key to our results and relies on the ex ante Pareto efficiency of the social choice function. An agent cannot gain, for instance, by reversing his announced types across two problems because the social choice function is already picking something that maximizes an agent's ex ante expected utility given the distribution of other agents' types; otherwise it would not have been ex ante Pareto efficient.

#### 4. PARTICIPATION CONSTRAINTS

Theorem 1 holds for any ex ante efficient social choice functions that we target. As such,  $f$  can satisfy any number of auxiliary properties, such as participation constraints (also commonly referred to as individual rationality constraints), fairness, etc. The interest in participation constraints often arises in settings where agents have a choice of whether or not to participate in the mechanism, and this might occur after they already know their preferences. This, for instance, is often a constraint in any contracting setting, including the bargaining setting we considered in Example 2. If such participation decisions are relevant, then it is important to demonstrate that they will be satisfied by linking mechanisms all along the sequence—and not just in the limit—especially because in some settings the conflict between incentive compatibility and efficiency arises only when an interim participation constraint is in place.



Consider a decision problem  $(D, U, P)$ , where some decision  $e \in D$  has a special designation, which may be thought of as a status quo, an endowment, or an outside option. The interpretation is that  $e$  will be the default decision if some agent(s) choose not to participate.

A social choice function  $f$  satisfies an *ex ante participation constraint* if  $E[u_i(f(u))] \geq E[u_i(e)]$  for all  $i$ , an *interim participation constraint* if  $E[u_i(f(u))|u_i] \geq u_i(e)$  for all  $i$  and  $u_i$ , and an *ex post participation constraint* if  $u_i(f(u)) \geq u_i(e)$  for all  $i$  and  $u$ . We say that  $f$  satisfies a strict ex ante, interim, or ex post participation constraint, respectively, if, in addition, for each  $i$  the respective constraint holds strictly (for at least one  $u_i$  in the interim case and at least one  $u$  in the ex post case).

Let us discuss why one needs some modification of the linking mechanism to be able to satisfy a participation constraint. Reconsider Example 2. Suppose that we have linked 500 problems and by chance the seller happens to be of a type that is at least 0.7 on all of the problems. By participating in the mechanism (under any subsequent equilibrium play) with this type, he has a negative expected utility. Thus, to satisfy the interim participation constraint (or the ex post constraint), we need to modify the linking mechanism.

Consider a decision problem  $(D, U, P)$  with an option of not participating that results in a status quo option, denoted  $e$ . Consider the following variation on the mechanism  $(M^K, g^K)$  that is used in the proof of Theorem 1. In a first stage, the agents submit their announcements from  $M_i^K$  and decisions on all problems are given by  $g^K(m^K)$ . In a second stage, agents are each asked (say simultaneously) whether they wish to participate or not. If any agent chooses not to participate, then  $e$  is selected on all problems; otherwise the outcomes are  $g^K(m^K)$ . We say that a strategy for an agent is approximately truthful if  $m_i$  is approximately truthful and an agent chooses not to participate only in situations where his utility from nonparticipation (getting  $e$  on all problems) exceeds the utility of  $g^K(m^K)$ .

So, we have modified the linking mechanism to explicitly allow agents an option to not participate. We have done this at the ex post stage, which will provide for the strongest of the three forms of a participation constraint. It is important to note, however, that an agent must decide to participate in the whole linking mechanism or not to participate at all. We discuss allowing an agent to pick some problems not others in which to participate in the Appendix.

**COROLLARY 1:** *Consider any ex ante efficient  $f$  that satisfies a strict participation constraint of any sort: ex ante, interim, or ex post. Consider the two-stage linking mechanisms with a participation decision as described previously. For every  $K$ , there exists an approximately truthful perfect Bayesian equilibrium of the modified two-stage linking mechanism such that the resulting social choice function sat-*

ifies an ex post (and thus interim and ex ante) participation constraint and the sequence of these equilibria approximate  $f$ .<sup>10</sup>

## 5. REMARKS

We have discussed our linking mechanisms as if all of the decisions were to be taken at the same time. When decisions are implemented across time, discounting is the effective determinant of the number of linked problems  $K$ .

To link  $K$  problems over time, consider the obvious variation of our previous mechanism made to operate over time. An agent is budgeted to announce a type of  $u_i$  exactly  $P_i(u_i)K$  times. Once an agent has used up his announcements of a given type, he cannot announce that type in any of the remaining periods. A stream of decisions  $(d^1, \dots, d^K)$  results in a utility of  $\sum_k \delta_i^k u_i^k(d^k)$  for agent  $i$ , where  $\delta_i \in (0, 1]$  is a discount factor. The following corollary follows from the security part of Theorem 1, with some special attention needed to treat the case of  $n \geq 3$ .<sup>11</sup>

**COROLLARY 2:** *Consider a decision problem  $(D, U, P)$  and an ex ante efficient social choice function  $f$  with corresponding ex ante expected utility levels  $(\bar{u}_1, \dots, \bar{u}_n)$ . For any  $\varepsilon > 0$ , there exists  $\bar{K}$  such that, for each  $K \geq \bar{K}$ , there exists  $\bar{\delta}$  such that, for every  $\delta \geq \bar{\delta}$ , every Bayesian equilibrium of the mechanism operating over time leads to an ex ante expected utility for each agent  $i$  that is above  $\bar{u}_i - \varepsilon$  (per problem).*

### *How Large Is Large?*

We can put a bound on the number of problems where any mistake will be made in the linking mechanism we have proposed here. A theorem of Kolmogorov ((13.4) in Billingsley (1968)) implies that the proportion of problems out of  $K$  on which agents might be forced to lie is on the order of  $\frac{1}{\sqrt{K}}$ . Whereas the secure strategies of approximate truth have a proportion of lies that are bounded by this, the percentage distance from full ex ante efficiency is on the order of  $\frac{1}{\sqrt{K}}$ . In many problems it is in fact closer, and we think that this is a fruitful subject for further research (e.g., see Cohn (2004)).<sup>12</sup>

*Dept. of Economics, Stanford University, Stanford, CA 94305, U.S.A.; jacksonm.@stanford.edu; <http://www.stanford.edu/~jacksonm/>*

<sup>10</sup>Approximate truth, coupled with a decision to participate whenever one has a higher utility from participating than the outside option, is still secure in the sense that it secures a utility level that approaches the ex ante efficiency when other agents participate.

<sup>11</sup>The modification of the mechanism to deal with collusion when  $n \geq 3$  (see the proof of Theorem 1) is somewhat subtle, but can be handled.

<sup>12</sup>See also Veszteg (2005), as well as Qu erou (2005), who looked at bargaining problems across different populations of players.

and

*Dept. of Economics, University of Chicago, 1126 E. 59th Street, Chicago, IL 60637, U.S.A.; h-sonnenschein@uchicago.edu.*

APPENDIX: PROOFS

PROOF OF THEOREM 1: Consider any  $K$  and the linking mechanism  $(M^K, g^K)$ . A strategy  $\sigma_i: U_i^K \rightarrow \Delta(M^K)$  for  $i$  is *label-free* if it depends only on the realization of  $i$ 's preferences and not the labels of the problems. Formally, given a permutation (bijection)  $\pi: \{1, \dots, K\} \rightarrow \{1, \dots, K\}$  and any  $u_i = (u_i^1, \dots, u_i^K) \in U_i^K$ , let  $u_i^\pi$  be defined by  $(u_i^\pi)^k = u_i^{\pi(k)}$  for each  $k \in \{1, \dots, K\}$ . Given our definition of  $M_i^K$  there is a corresponding notion of  $m_i^\pi$  that starts from any  $m_i \in M_i^K$ . A strategy  $\sigma_i$  for  $i$  is *label-free* if for any permutation  $\pi: \{1, \dots, K\} \rightarrow \{1, \dots, K\}$ ,  $\sigma_i(u_i^\pi)[m_i^\pi] = \sigma_i(u_i)[m_i]$ , where  $\sigma_i(u_i)[m_i]$  is the probability of playing  $m_i$  at  $u_i$  under  $\sigma_i$ .

The modification of the linking mechanism  $(M^K, g^K)$  for more than two agents is as follows: For any subset of agents  $C$ ,  $m_C \in M_C^K$ , and set of problems  $T \subset \{1, \dots, K\}$ , let  $F_C^K(m_C, T) \in \Delta(U_C)$  be the frequency distribution of announced profiles of types by  $C$  on problems in  $T$ . Thus, this is a distribution on  $U_C$  conditional on looking only at the announcements made on problems in  $T$ . For any agent  $i$ , any coalition  $C$  such that  $i \notin C$ , and any announced vector of  $m \in M^K$ , consider the measure

$$d_{i,C}^K(m) = \max_{\substack{u_C \in U_C \\ u_i \in U_i}} |P_C^K[u_C] - F_C^K(m_C, \{k | m_i^k = u_i\})[u_C]|,$$

where  $P_C^K = \prod_{j \in C} P_j^K$ . If this measure differs significantly from 0, then the group  $C$ 's announcements differ significantly from the underlying distribution. That is, this measure looks at the distribution of the announced  $u_C$ 's conditional on the dates that  $i$  announced some  $u_i$  and checks whether it is close to what the empirical distribution should be.

Given a sequence of numbers  $\varepsilon^K > 0$  to be described shortly, modify the mechanism  $(M^K, g^K)$  as follows. Consider an announcement  $m \in M^K$ . We redefine  $g^K(m)$  as follows: For each  $i$ , identify a smallest  $C$  (breaking ties in any fixed manner) for which  $d_{i,C}^K(m) > \varepsilon^K$ , if any exists.<sup>13</sup> Starting with the lowest index  $i$  for which there is such a  $C$  (if any), instead of using  $m_C$ , generate a random announcement  $\tilde{m}_C$  to replace it, by independently picking a message  $\tilde{m}_j \in M_j^K$  (with equal probability on each message) for each  $j \in C$  and then substitute  $\tilde{m}_C$  for  $m_C$ . Now keep iterating on this process until there is

<sup>13</sup>It is important to look for the smallest such subsets, because otherwise we might end up penalizing "honest" agents along with manipulating coalitions, which would skew incentives.

no  $i$  and  $C$  for which  $d_{i,C}^K(m') > \varepsilon^K$ , where  $m'$  is the announcement that includes all modifications from previous steps of the process. The mechanism then uses the final announcement from this process and so generates  $g^K(m')$  as an outcome. By a strong law of large numbers of distributions, such as the Glivenko–Cantelli theorem (see Billingsley (1968)), we can find  $\varepsilon^K \rightarrow_K 0$ , such that for any strategies  $\sigma$ , if any agent  $j$ 's strategies are approximately truthful, then the probability that  $j$ 's announcements are modified under this process vanishes.

We make one further modification of the mechanism. For a given  $K$ , the distribution  $P_i^K$  may not exactly match  $P_i$ . To make sure that for an arbitrary decision problem we always have an approximately truthful equilibrium, we need to be sure that the distributions exactly match  $P_i$  and not just approximately.<sup>14</sup> The following modification of the linking mechanisms ensures this. Find the smallest  $\gamma^K \geq 0$  such that there exists another distribution  $\tilde{P}_i^K$  such that  $(1 - \gamma^K)P_i^K + \gamma^K\tilde{P}_i^K = P_i$ . Note that  $\gamma^K \rightarrow 0$ . On any given problem  $k$ , let the mechanism  $g^K$  follow  $i$ 's announced (possibly modified by the preceding procedure)  $m_i^k$  with probability  $(1 - \gamma^K)$ , and with probability  $\gamma^K$ , randomly draw an announcement according to  $\tilde{P}_i^K$ , doing this independently across problems and agents.

We first prove (iii). Consider the following “approximately truthful” strategy  $\sigma_i^K$ . Consider a realized  $u_i \in U_i^K$ . For any  $v_i \in U_i$  with frequency less than  $P_i^K(v_i)$  in the vector  $u_i$ , announce truthfully on all problems  $k$  such that  $u_i^k = v_i$ . For other  $v_i$ 's, with equal probability pick  $K \times P_i^K(v_i)$  of the problems  $k$  such that  $u_i^k = v_i$  on which to announce truthfully. On the remaining problems randomly pick announcements to satisfy the constraints imposed by  $P_i^K$  under  $M_i^K$  and place equal probability on all such announcements. Given the independence of types, this is label-free and independent of the announcements of all other agents on all problems. By using  $\sigma_i^*$ , agent  $i$  guarantees himself an expected utility  $\bar{u}_i^K$  per problem that approaches the utility that comes under truth-telling by all agents, regardless of the strategy of the other agents. This follows because by construction of the mechanism, the agent is guaranteed that, conditional on the problems where the agent announces any given  $v_i$ , the distribution over other agents' types is approximately independently distributed and approximately what should be expected if the other agents were truthful (regardless of whether they are), and that the chance that the agent's strategy will be replaced by an  $\tilde{m}_i$  is vanishing. This implies that the sequence  $\bar{u}_i^K$  converges to  $\bar{u}_i = E[u_i(f(u))]$ . Moreover, the same conclu-

<sup>14</sup>For some decision problems, it might be that  $f$  is ex ante efficient for the given  $P_i$ , but not quite for some approximations of it. This ex ante Pareto efficiency of  $f$  relative to an agent's expectations plays an important role in obtaining an approximately truthful equilibrium. Note, however, that for two-player settings this modification is not needed to establish that all equilibria converge to being efficient; it is only needed to establish the existence of approximately truthful equilibria.

sion follows for any approximately truthful strategy,<sup>15</sup> and so we have established (iii).

We next establish (iv). Whereas every agent can obtain an expected utility per problem of at least  $\bar{u}_i$  in the limit—regardless of the other agents' strategies—by following the “approximately truthful” strategy  $\sigma_i^K$ , then it must be that the lim inf of each agent's expected utility per problem along any sequence of equilibria is at least  $\bar{u}_i$ . Next, note that by the ex ante Pareto efficiency of  $f$ , for any profile of strategies and any  $K$ , if some agent  $i$  is expecting a utility higher than  $\bar{u}_i$ , then some other agent  $j$  must be expecting a utility of less than  $\bar{u}_j$ . However, because the lim inf of each agent's expected utility for any sequence of equilibria is at least  $\bar{u}_i$ , then it must be that this is the limit of the expected utility of each agent, and thus every equilibrium's expected utility profile must converge to the desired limit.

We now argue (v). Consider any sequence of equilibria  $\sigma^K$  and some sequence of deviating coalitions  $C^K$ . Consider any agent  $i$  who appears in the sequence  $C^K$  infinitely often (other agents are of no consequence to the conclusion). As previously mentioned, regardless of  $i$ 's strategies, for large enough  $K$ , conditional on the problems where agent  $i$  announces any given type  $u_i$ , the distribution over other agents' eventual types under the mechanism are approximately independently distributed and approximately what should be expected if the other agents were truthful (regardless of whether they are). Given the private values in the model, it is without consequence to  $i$ 's utility whether the other agents' announcements as evaluated by the mechanism (after any modifications as described) are truthful or not, so without loss of generality for  $i$ 's utility we can take the other agents' announcements to be as if they were their true utility functions. Thus, the profile of utilities for the other agents on the problems where  $i$  announces some type  $u_i$  converge to what they should (under truth and  $f$ ) conditional on  $i$  being of type  $u_i$ , on problems where  $i$  announces  $u_i$ . Taking a weighted average across  $i$ 's types, this means that the average utility for each other agent  $j$  across problems is converging to  $\bar{u}_j$ . By the ex ante Pareto efficiency of  $f$ , this implies that  $i$ 's expected utility per problem cannot converge to be more than  $\bar{u}_i$ .

To conclude the proof, let us show (i) and (ii). We show a strong version of (i), namely that there is a label-free approximately truthful equilibrium. Then (ii) follows from a law of large numbers. Consider any agent  $i$ . If all agents  $j \neq i$  play label-free strategies, then given the definition of the strategy spaces  $M_j^K$  and the independence across problems, the distribution of the announcements of agents  $j \neq i$  on any problem is given by  $P_{-i}$ , and this is iden-

<sup>15</sup>Although such a strategy might not be completely independent of other player's strategies, it still follows from the construction of the mechanism that as  $K$  grows, conditional on the problems where the agent announces any given type  $v_i$ , other agents' types are arbitrarily close to being independently distributed and approximately what should be expected if the other agents were truthful, and that the chance that the agent's strategy will be replaced by an  $\tilde{m}_i$  is vanishing.

tically distributed across problems.<sup>16</sup> It then follows that for any best response that  $i$  has to label-free strategies of the other agents, there will be a label-free best response for  $i$ .<sup>17</sup> Note also that any best response to some label-free strategies of other agents is a best response to *any* label-free strategies of the other agents, because, given the independence, all label-free strategies of the other agents induce the same distribution over announcements by the other agents on any set of problems. Given the finite nature of the game, for any set of label-free strategies of agents  $-i$  there exists a best response for agent  $i$  and, as argued in the preceding text, one that is label-free. Thus there exists a label-free equilibrium.

Next, let us show that there exists such an equilibrium that is approximately truthful in the sense that  $i$  never permutes the announcements of his true utility functions across some set of problems.

Consider any  $K$  and a label-free equilibrium  $\sigma$ . Consider some  $m_i = (\hat{u}_i^1, \dots, \hat{u}_i^K) \in M_i^K$  such that  $\sigma_i(u_i)[m_i] > 0$  for some  $(u_i^1, \dots, u_i^K) \in U_i^K$ . Suppose that there is some subset of problems  $T \subset \{1, \dots, K\}$  such that  $i$  is permuting announcements on  $T$  under  $m_i$ . That is there exists a permutation  $\pi: T \rightarrow T$  such that  $\pi(k) \neq k$  and  $\hat{u}_i^k = u_i^{\pi(k)}$  for all  $k \in T$ . So  $i$ 's announcement under  $m_i$  reshuffles the true utility functions that  $i$  has under  $u_i$  on the problems in  $T$  according to  $\pi$ . Consider replacing  $m_i$  with  $\tilde{m}_i$ , where this permutation on  $T$  is replaced by truthful announcing. That is,  $\tilde{m}_i^k = u_i^k$  for each  $k \in T$  and  $\tilde{m}_i^k = m_i^k$  for each  $k \notin T$ . Then consider an alternative strategy, denoted  $\tilde{\sigma}_i$ , that differs from  $\sigma_i$  only at  $u_i$  and then sets  $\tilde{\sigma}_i(u_i)[m_i] = 0$  and  $\tilde{\sigma}_i(u_i)[\tilde{m}_i] = \sigma_i(u_i)[\tilde{m}_i] + \sigma_i(u_i)[m_i]$ . The claim is that  $\tilde{\sigma}_i$  leads to at least as high an expected utility as  $\sigma_i$ . This follows from the ex ante Pareto efficiency of  $f$ . To see this, note that the distribution of announcements under either strategy together with the strategies of the other agents is  $P$  on each problem and is independent across all problems (given the label-free nature of the strategies). Thus, the other agents' ex ante expected utilities on any given problem are not affected by the change in strategies. If  $i$ 's utility were to fall as a result of using  $\tilde{\sigma}_i$  instead of  $\sigma_i$ , then it would be that  $f$  could be Pareto improved upon by a corresponding change to some  $f'$  that took  $u_i$ 's and remapped them as done under  $\pi$  (with corresponding probabilistic weights). This would contradict the ex ante Pareto efficiency of  $f$ . Note that we can do this in a way that preserves the label-free nature of the strategy by randomly picking one from all sets  $T$  where the permutation  $\pi$  is used. Now we can continue to undo such permutations until we have reached a label-free strategy that has no such permutations. This is the "approximately truthful" strategy that we

<sup>16</sup>Announcements are not independent across problems, because the constraints imposed by  $M^K$  prevent this.

<sup>17</sup>Starting with any best response that is label dependent, given that other agents, strategies are label-free, any variation based on permuting the dependence on labels will also be a best response, as will a convex combination of such permutations, which is label-free.

sought, and it still provides at least the utility of  $\sigma_i$  and thus is a best response. In addition, because it is label-free, it follows that the overall equilibrium is still preserved. Iterating on agents leads to the desired profile of equilibrium strategies. *Q.E.D.*

SKETCH OF THE PROOF OF COROLLARY 1: Have agents play the approximately truthful and label-free equilibrium strategy identified in the proof of Theorem 1 in the first stage of the mechanism. Let  $m^K$  be the announcements from the first stage. If agent  $i$ 's utility<sup>18</sup>  $\sum_k u_i^k(f(m^k))$  is at least  $\sum_k u_i^k(e)$ , then have  $i$  agree to participate; otherwise, have the agent choose not to participate. Given the "label-free" nature of the first stage strategies (see the proof of Theorem 1 for details), the choices of agents to participate are independent and do not affect the equilibrium structure of the first stage announcements. By a strong law of large numbers (e.g., the Glivenko–Cantelli theorem), as  $K$  becomes large,  $(\sum_k u_i^k(f(m^k)))/(KE[u_i(f(u))])$  converges to 1 in probability, as does  $(\sum_k u_i^k(e))/(KE[u_i(e)])$ . Given that some version of a strict participation constraint is satisfied by  $f$ , it follows that  $E[u_i(f(u))] > E[u_i(e)]$ . Thus,  $\sum_k [u_i^k(f(m^k))] > \sum_k u_i^k(e)$  with probability approaching 1.

We remark that the corollary would not go through if we allowed agents to choose whether to participate problem by problem. For some settings, an agent can manipulate such a mechanism by lying selectively in stage one and then selectively opting out in the second stage. Nevertheless, we can modify the mechanism so that by being approximately truthful an agent guarantees himself a probability approaching 1 of satisfying an ex post constraint on every problem, while getting the ex ante expected utility level. We do this by loosening the budget constraint so that agents can slightly overannounce any of their types in such a way that the probability that any agent has to lie goes to 0, while the budget constraint still converges to the true distribution of types. Such a mechanism has label-free equilibria that approach truthful announcements, agents satisfy an interim constraint on every problem with a probability approaching 1, and all agents satisfy an ex post constraint on a fraction of problems approaching 1 with a probability approaching 1.

#### REFERENCES

- ABREU, D., D. PEARCE, AND E. STACCHETTI (1991): "Toward a Theory of Discounted Repeated Games with Imperfect Monitoring," *Econometrica*, 58, 1041–1065. [242]  
 ADAMS, W. J., AND J. L. YELLEN (1976): "Commodity Bundling and the Burden of Monopoly," *Quarterly Journal of Economics*, 90, 475–498. [242]  
 ARMSTRONG, M. (1999): "Price Discrimination by a Many-Product Firm," *Review of Economic Studies*, 66, 151–168. [242]

<sup>18</sup>We are abusing notation slightly because  $f$  may be randomizing on outcomes, in which case we are looking at an expected utility.



- BAKOS, Y., AND E. BRYNJOLFSSON (1999): "Bundling Information Goods: Pricing, Profits, and Efficiency," *Management Science*, 45, 1613–1630. [242]
- (2000): "Bundling and Competition on the Internet," *Marketing Science*, 19, 63–82. [242]
- BERNHEIM, B. D., AND M. D. WHINSTON (1990): "Multimarket Contact and Collusive Behavior," *Rand Journal of Economics*, 21, 1–26. [243]
- BILLINGSLEY, P. (1968): *Convergence of Probability Measures*. New York: Wiley. [250,252]
- BLUME, A., AND P. HEIDHUES (2002): "Tacit Collusion in Repeated Auctions," Mimeo, University of Pittsburgh. [243]
- CAMPBELL, C. M. (1998): "Coordination in Auctions with Entry," *Journal of Economic Theory*, 82, 425–450. [243]
- CASELLA, A. (2002): "Storable Votes," Mimeo, Columbia University. [243]
- CASELLA, A., AND T. R. PALFREY (2003): "Storable Votes," Mimeo, Columbia University and California Institute of Technology. [243]
- CHAKRABORTY, A., N. GUPTA, AND R. HARBAUGH (2002): "Ordinal Cheap Talk in Common Value Auctions," Mimeo, Baruch College CUNY, University of Michigan, and Claremont McKenna College. [243]
- CHAKRABORTY, A., AND R. HARBAUGH (2003): "Cheap-Talk Comparisons in Multi-Issue Bargaining," *Economics Letters*, 78, 357–363. [242]
- COHN, Z. (2004): "A Note on Linked Bargaining," Mimeo, University of Chicago. [250]
- CRÉMER, J., AND R. MCLEAN (1988): "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56, 1247–1257. [245]
- FANG, H., AND P. NORMAN (2003): "An Efficiency Rational for Bundling of Public Goods," Discussion Paper 1441, Cowles Foundation. [242]
- FUDENBERG, D., AND D. LEVINE (1994): "Efficiency and Observability in Games with Long-Run and Short-Run Players," *Journal of Economic Theory*, 62, 103–135. [242]
- GREEN, E. J., AND R. PORTER (1984): "Noncooperative Collusion under Imperfect Price Information," *Econometrica*, 52, 87–100. [242]
- HÖRNER, J., AND J. S. JAMISON (2004): "Collusion with (Almost) No Information," Mimeo, MEDS. [242]
- HORTALA-VALLVE, R. (2003): "Qualitative Voting," Mimeo, LSE. [243]
- MASKIN, E., AND J. TIROLE (1990): "The Principal–Agent Relationship with an Informed Principal: The Case of Private Values," *Econometrica*, 58, 379–409. [242]
- MAY, K. (1952): "A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decision," *Econometrica*, 20, 680–684. [244]
- MCAFEE, R. P. (1992): "Amicable Divorce: Dissolving a Partnership with Simple Mechanisms," *Journal of Economic Theory*, 56, 266–293. [242,243]
- MCAFEE, R. P., J. MCMILLAN, AND M. D. WHINSTON (1979): "Multiproduct Monopoly, Commodity Bundling and the Correlation of Values," *Quarterly Journal of Economics*, 100, 371–383. [242]
- MILLER, N. R. (1977): "Logrolling, Vote Trading, and the Paradox of Voting: A Game-Theoretical Overview," *Public Choice*, 30, 51–75. [242]
- MYERSON, R., AND M. SATTERTHWAITTE (1983): "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, 29, 265–281. [245]
- PESENDORFER, M. (2000): "A Study of Collusion in First-Price Auctions," *Review of Economic Studies*, 67, 381–412. [243]
- QUÉROU, N. (2005): "Obtaining a Unique, Efficient Norm Through Negotiation by Linking Proposals," Mimeo, CREUSET, Saint-Étienne. [250]
- RADNER, R. (1981): "Monitoring Cooperative Agreements in a Repeated Principal–Agent Relationship," *Econometrica*, 49, 1127–1148. [242]
- TOWNSEND, R. (1982): "Optimal Multi-Period Contracts and the Gain from Enduring Relationships under Private Information," *Journal of Political Economy*, 90, 1166–1186. [242]
- TULLOCK, G. (1970): "A Simple Algebraic Logrolling Model," *American Economic Review*, 60, 419–426. [242]



- VESZTEG, R. (2005): "Linking Decisions with Moments," Mimeo, Universidad de Navarra. [250]
- WILSON, R. (1969): "An Axiomatic Model of Logrolling," *American Economic Review*, 59, 331–341. [242]