COMMENT ON JACKSON AND SONNENSCHEIN (2007) "OVERCOMING INCENTIVE CONSTRAINTS BY LINKING DECISIONS"

IAN BALL
Department of Economics, MIT

MATTHEW O. JACKSON
Department of Economics, Stanford University

DENIZ KATTWINKEL
Department of Economics, UCL

WE CORRECT A BOUND in the definition of approximate truthfulness used in the body of the paper of Jackson and Sonnenschein (2007). The proof of their main theorem uses a different permutation-based definition, implicitly claiming that the permutation-version implies the bound-based version. We show that this claim holds only if the bound is loosened. The new bound is still strong enough to guarantee that the fraction of lies vanishes as the number of problems grows, so the theorem is correct as stated once the bound is loosened.

Setting. Recall the setting of Jackson and Sonnenschein (2007, hereafter JS). Consider an n-agent collective decision problem $\mathcal{D} = (D, U, P)$, where D is the finite set of decisions; $U = U_1 \times \cdots \times U_n$ is the finite set of possible profiles of utility functions on D; and $P = (P_1, \ldots, P_n)$ in $\Delta(U_1) \times \cdots \times \Delta(U_n)$ is the profile of priors.

There are K independent copies of this decision problem, labeled $k = 1, \ldots, K$. Each agent i knows their preference vector $u_i = (u_i^1, \ldots, u_i^K)$ in U_i^K, and their total payoff from a decision vector $(d^1, \ldots, d^K)$ is the sum $u_i^1(d^1) + \cdots + u_i^K(d^K)$. Utility functions are drawn independently across agents and decision problems, according to the priors in P.

Given an ex ante Pareto efficient social choice function $f : U \to \Delta(D)$, JS introduce the following *linking mechanism*. Each agent i is asked to report a preference vector $\hat{u}_i = (\hat{u}_i^1, \ldots, \hat{u}_i^K)$ with exactly the same utility frequencies as the distribution P_i^K, where P_i^K is the closest approximation of P_i with the property that every probability is a multiple of $1/K$. In each decision problem, the mechanism applies the social choice function f to the profile of preferences reported on that problem.

To formalize the linking mechanism, define the *marginal (distribution)* of a vector u_i in U_i^K, denoted either $\text{marg}\, u_i$ or $\text{marg}(\cdot | u_i)$, by

$$\text{marg}(v_i | u_i) = \#\{k : u_i^k = v_i\}/K, \quad v_i \in U_i.$$

The linking mechanism is a pair (M^K, g^K). The message space M^K equals the product $M_1^K \times \cdots \times M_n^K$, where

$$M_i^K = \{\hat{u}_i \in U_i^K : \text{marg}\, \hat{u}_i = P_i^K\}.$$

Ian Ball: ianball@mit.edu
Matthew O. Jackson: jacksonm@stanford.edu
Deniz Kattwinkel: d.kattwinkel@ucl.ac.uk
We thank Yifan Dai for excellent research assistance.

The outcome rule $g^K : M^K \to (\Delta(D))^K \subset \Delta(D^K)$ is defined by[1]

$$g^K(\hat{u}^1, \ldots, \hat{u}^K) = (f(\hat{u}^1), \ldots, f(\hat{u}^K)),$$

where $\hat{u}^k$ denotes the vector $(\hat{u}_1^k, \ldots, \hat{u}_n^k)$ of reports on problem k.

In the linking mechanism, an agent cannot report exactly truthfully if their realized preference vector violates the marginal constraint. JS focus on strategies in which each agent lies in as few decision problems as is feasible under the mechanism. They define a strategy $\sigma_i^K : U_i^K \to M_i^K$ to be *approximately truthful* if

$$\#\{k : [\sigma_i^K(u_i)]^k \neq u_i^k\} = \min_{\hat{u}_i \in M_i^K} \#\{k : \hat{u}_i^k \neq u_i^k\}, \tag{1}$$

for all u_i in U_i^K, where $[\sigma_i^K(u_i)]^k$ denotes the kth component of $\sigma_i^K(u_i)$. The right side of (1) equals $Kd(\text{marg}\, u_i, P_i^K)$, where d is the total variation metric on $\Delta(U_i)$ defined by $d(Q, Q') = \sum_{v_i \in U_i} (Q(v_i) - Q'(v_i))_+$.[2]

We propose a weaker bound. A strategy $\sigma_i^K : U_i^K \to M_i^K$ is *approximately truthful** if

$$\#\{k : [\sigma_i^K(u_i)]^k \neq u_i^k\} \leq (\#U_i - 1)Kd(\text{marg}\, u_i, P_i^K), \tag{2}$$

for all u_i in U_i^K. The inequality in (2) relaxes (1) by a factor of $(\#U_i - 1)$. Agent i can lie $(\#U_i - 1)$ times more than is required by the marginal constraint. The two definitions coincide if $\#U_i \leq 2$. For $\#U_i > 2$, the new definition is strictly weaker. Both definitions extend immediately to mixed strategies.[3]

We also give a name to a different notion of truthfulness that appears in JS's proof. A strategy $\sigma_i^K : U_i^K \to \Delta(M_i^K)$ is *permutation-truthful* if, for each u_i in U_i^K and each $\hat{u}_i$ in $\text{supp}\, \sigma_i^K(u_i)$, the following holds: for any subset S of $\{1, \ldots, K\}$, if there is a bijection π on S such that $\hat{u}_i^k = u_i^{\pi(k)}$ for all k in S, then $\hat{u}_i^k = u_i^k$ for all k in S. That is, an agent never nontrivially permutes their true preferences over a subset of decision problems.

The concepts of permutation- and approximate truthfulness serve distinct roles in JS's argument. Permutation-truthfulness captures the reporting incentives created by the efficiency of the outcome function. Approximate truthfulness directly quantifies how frequently the reports match the truth.

Theorem 1.i in JS (p. 248) says that each linking mechanism has a Bayesian equilibrium in approximately truthful strategies; however, their proof shows only that there exists a Bayesian equilibrium in (label-free[4]) permutation-truthful strategies. We give a counterexample to the existence of an approximately truthful equilibrium. Next, we prove that permutation-truthful strategies are approximately truthful* and that this weaker property is still sufficient for part (ii) of Theorem 1. The remaining parts of the theorem are true as stated—only part (iii) mentions approximate truthfulness and it is true under either definition. Therefore, Theorem 1 is correct if *approximately truthful* is everywhere replaced by *approximately truthful**.

---

[1]This exact outcome rule is used only in the case $n = 1$. For $n > 1$, the reports are modified before applying f in such a way that the modified reports follow P exactly.

[2]The right side of (1) is the minimal cost in the optimal transport problem between measures $K \text{marg}\, u_i$ and KP_i^K with cost function $c(x, y) = [x \neq y]$.

[3]A mixed strategy is approximately truthful (approximately truthful*) if it can be expressed as a mixture over approximately truthful (approximately truthful*) pure strategies.

[4]This means that whenever the agent's preference vector u_i is permuted, their report $\sigma_i^K(u_i)$ is permuted in the same way; see JS (p. 251).

TABLE I

COUNTEREXAMPLE.

	1	2	3
type vector	A	A	B
approximate truth	A	C	B
deviation	A	B	C

Counterexample.   Suppose there is a single agent ($n = 1$). The set of decisions is $D = \{a, b, c\}$. The agent has three possible utility functions, denoted $u(\cdot|A)$, $u(\cdot|B)$, and $u(\cdot|C)$. The prior P puts probability $1/3$ on each utility function. We say that the agent's *type* is either A, B, or C. Suppose that type A (respectively B, C) strictly prefers decision a (respectively b, c) to the other two decisions. Therefore, the unique ex ante Pareto efficient social choice function is $f(A) = a$, $f(B) = b$, and $f(C) = c$.

Consider linking $K = 3$ decisions. In the linking mechanism, the agent must report a vector $\hat{u} = (\hat{u}^1, \hat{u}^2, \hat{u}^3)$ in which A, B, and C each appear exactly once. Suppose the agent has type vector (A, A, B), as indicated in Table I. Reporting truthfully would violate the quota. Under an approximately truthful strategy, the agent must lie exactly once by reporting either (A, C, B) or (C, A, B). But the agent strictly prefers to lie twice by reporting (A, B, C) if

$$u(b|A) + u(c|B) > u(c|A) + u(b|B),$$

which holds as long as $u(c|A)$ is low enough. In this case, reporting C in problem 2 is so costly that the agent prefers to report B even though this forces them to lie again in problem 3 to satisfy the quota.[5] Reporting (A, B, C) does not violate approximate truthfulness* or permutation-truthfulness.

Approximate Efficiency.   Theorem 1.ii in JS says that their sequence $\{\sigma^K\}$ of equilibria *approximate* f in the sense that

$$\lim_K \left[\max_{k \leq K} \mathbf{P}\{g_k^K(\sigma^K(u)) \neq f(u^k)\} \right] = 0, \tag{3}$$

where g_k^K denotes the component of g^K in the kth decision problem; the equation inside the probability is between lotteries in $\Delta(D)$; and the probability is taken over the random vector u in U^K and possible mixing in σ^K. By the definition of g^K, (3) holds as long as we have

$$\lim_K \left[\max_{k \leq K} \mathbf{P}([\sigma^K(u)]^k \neq u^k) \right] = 0. \tag{4}$$

JS observe that (4) follows from the law of large numbers for label-free approximately truthful strategy profiles. We confirm that (4) holds for label-free approximately truthful* strategy profiles $\{\sigma^K\}$. For such profiles, we have for each agent i that

$$\max_{k \leq K} \mathbf{P}([\sigma_i^K(u_i)]^k \neq u_i^k) \leq (\#U_i - 1)\, \mathbf{E}[d(\operatorname{marg} u_i, P_i^K)]$$

$$\leq (\#U_i - 1)\, \mathbf{E}[d(\operatorname{marg} u_i, P_i)] + (\#U_i - 1)d(P_i, P_i^K).$$

---

[5]In general, telling a different lie in one problem can start a cycle of up to $\#U_i - 1$ lies in total. If the cycle had length $\#U_i$, then the agent could report truthfully on each problem in the cycle.

As $K \to \infty$, the approximation P_i^K converges to P_i, and the expectation goes to zero by the law of large numbers (since U_i is finite). Then (4) follows by applying a union bound over the agents $i = 1, \ldots, n$.[6]

To complete the proof, we check that permutation-truthfulness implies approximate truthfulness*. Consider an agent i. The following combinatorial lemma guarantees that there is a sufficiently large subset S of decision problems over which the agent permutes their true preferences. Under a permutation-truthful strategy, the agent is truthful on S, so the size of the complement of S gives the desired bound (2) on the number of lies.

LEMMA 1: *Fix an agent i. For any pair of vectors u_i and $\hat{u}_i$ in U_i^K, there exists a subset S of $\{1, \ldots, K\}$ and a bijection π on S such that*
  (i) $\hat{u}_i^k = u_i^{\pi(k)}$ *for all k in S;*
  (ii) $\#S \geq K - (\#U_i - 1) K d(\operatorname{marg} u_i, \operatorname{marg} \hat{u}_i)$.

PROOF: Given u_i and $\hat{u}_i$ in U_i^K, construct a directed multigraph as follows. The set of nodes is U_i. For each $k = 1, \ldots, K$, add edge k from node u_i^k to node $\hat{u}_i^k$. In this graph, each node v_i has out-degree $\deg^+(v_i) = K \operatorname{marg}(v_i|u_i)$ and in-degree $\deg^-(v_i) = K \operatorname{marg}(v_i|\hat{u}_i)$.

Now we add *new* edges as follows. Add an edge from a node with net in-degree to a node with net out-degree, update the degrees of the new graph, and repeat until the graph is balanced, that is, $\deg^+(v_i) = \deg^-(v_i)$ for all v_i in U_i. Let K' be the number of new edges added by this procedure. Counting the heads of the new edges, we have

$$K' = \sum_{v_i} \left(\deg^+(v_i) - \deg^-(v_i) \right)_+$$

$$= K \sum_{v_i} \left(\operatorname{marg}(v_i|u_i) - \operatorname{marg}(v_i|\hat{u}_i) \right)_+$$

$$= K d(\operatorname{marg} u_i, \operatorname{marg} \hat{u}_i).$$

Now we have a balanced graph with $K + K'$ edges. Partition this graph into edge-disjoint cycles.[7] Remove every cycle that contains at least one of the new edges. Define S to be the set of labels of the remaining edges. Since at most $K' \#U_i$ edges were removed, we have

$$\#S \geq K + K' - K' \#U_i = K - (\#U_i - 1) K d(\operatorname{marg} u_i, \operatorname{marg} \hat{u}_i),$$

so S satisfies (ii). For (i), define π on S by letting $\pi(k)$ be the label of the edge that follows edge k in its cycle. (In particular, $\pi(k) = k$ if edge k is a loop.) Since the head of edge k equals the tail of edge $\pi(k)$, we have $\hat{u}_i^k = u_i^{\pi(k)}$ by the definition of the graph.   Q.E.D.

REFERENCES

JACKSON, MATTHEW O., AND HUGO F. SONNENSCHEIN (2007): "Overcoming Incentive Constraints by Linking Decisions," *Econometrica*, 75, 241–257. [o3]

---

[6]For any (not necessarily label-free) approximately truthful* strategies σ_i^K, we can still conclude that the expected fraction of reports that are lies, $\mathbf{E}[\#\{k : [\sigma_i^K(u_i)]^k \neq u_i^k\}]/K$, tends to 0 as $K \to \infty$. This is the only property of approximately truthful strategies that is used in JS's proof of Theorem 1.iii-v.

[7]To do so, start at a node with an outgoing edge. Form a path by arbitrarily selecting outgoing edges until the path contains a cycle. Remove the cycle and repeat. Since the graph remains balanced, this process can terminate only when every edge has been removed.