Comment on Jackson and Sonnenschein (2007) "Overcoming Incentive Constraints by Linking Decisions"

Ian Ball* Matthew O. Jackson† Deniz Kattwinkel‡

January 19, 2022

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 is only true 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 and proof are 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 $f \colon U \to \Delta(D)$, JS introduce the following *linking mechanism*. Each agent i is asked to report a preference

*Department of Economics, Massachusetts Institute of Technology, ianball@mit.edu.
†Department of Economics, Stanford University, jacksonm@stanford.edu
‡Department of Economics, University College London, d.kattwinkel@ucl.ac.uk.

vector $\hat{u}_i = (\hat{u}_i^1, \ldots, \hat{u}_i^K)$ in which the reported preferences display exactly the same frequencies as the distribution P_i^K, which is the closest approximation of P_i such that $P_i^K(v_i)$ is a multiple of $1/K$ for each v_i in U_i. 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 $\operatorname{marg} u_i$ or $\operatorname{marg}(\cdot|u_i)$, by

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

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

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

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

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

In the linking mechanism, agents may be unable to report exactly truthfully because of the marginal constraint. JS focus on strategies in which each agent lies in as few decision problems as is feasible under the mechanism. In the body of their paper, they define a strategy $\sigma_i^K \colon U_i^K \to M_i^K$ for agent i to be *approximately truthful* if

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

for all $u_i \in U_i^K$ and $\hat{u}_i \in M_i^K$, where $[\sigma_i^K(u_i)]^k$ denotes the component of $\sigma_i^K(u_i)$ in the k-th decision problem.

We propose a weaker bound and say that a strategy $\sigma_i^K \colon U_i^K \to M_i^K$ for agent i is *approximately truthful** if

$$\#\{k : [\sigma_i^K(u_i)]^k \neq u_i^k\} \leq (\#U_i - 1) \sum_{v_i \in U_i} \left(\#\{k | u_i^k = v_i\} - P_i^K(v_i)K\right)_+,$$

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

for all $u_i \in U_i^K$. The summation captures the number of dimensions on which the agent must lie in order to satisfy the constraint—it equals the minimum of the right side of (1) over all reports $\hat{u}_i$ in M_i^K. This bound is weaker than the original definition by the factor $(\#U_i - 1)$, which reflects the total number of swaps that an agent may choose for each lie required by the mechanism. Both the original and revised definitions extend immediately to mixed strategies.[2]

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

The distinction between permutation and approximate truthfulness is useful because permutation-truthfulness naturally captures the incentives to report due to the efficiency of the outcome function, while approximate truthfulness directly quantifies how closely the reports match the expected frequency.

Theorem 1 in JS states that there exists an approximately truthful Bayesian equilibrium; however, their proof shows only that there is a permutation-truthful Bayesian equilibrium. We give a counterexample to the existence of an approximately truthful equilibrium, but then we show that the claim is correct with approximately truthful* in place of approximately truthful.

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 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). They cannot report truthfully since doing so would violate the reporting quota. Under an approximately truthful strategy,

[2] A mixed strategy is approximately truthful (approximately truthful*) if it is a mixture over approximately truthful (approximately truthful*) pure strategies.

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

Table 1. Counterexample

they must report (A, C, B) or (C, A, B). But they strictly prefer to report (A, B, C) if

$$u(b|A) + u(c|B) > u(c|A) + u(b|B). \tag{2}$$

which holds as long as $u(c|A)$ is low enough. Reporting (A, B, C) is permutation-truthful but not approximately truthful.

Table 1 displays these reports. It is feasible for the agent to lie only once by reporting C in problem 2. But under condition (2), the agent strictly prefers a different lie, namely, reporting B in problem 2, even though this report forces the agent to lie again in problem 3 in order to satisfy the reporting quota. This example does not violate the bound in approximate truthfulness*, which allows two lies instead of one. More generally, approximate truthfulness* allows for such a lying cascade of size $\#U_i - 1$ for every lie in an approximately truthful strategy.[3]

Approximate efficiency Theorem 1 in JS also states that the approximately truthful strategy profiles σ^K *approximate* f in the sense that

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

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

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

In their proof, JS note that (4) follows from the law of large numbers. The

[3] There cannot be a full cycle of $\#U_i$ lies for then the agent could report exactly truthfully.

same argument goes through for approximately truthful* strategy profiles σ^K, as we now confirm. If agent i has true preference vector u_i, then under an approximately truthful* strategy, they will lie at most $(\#U_i - 1)Kd(\operatorname{marg} u_i, P_i^K)$ times, where d denotes the total variation metric on $\Delta(U_i)$ defined by $d(Q, Q') = \sum_{v_i \in U_i}(Q(v_i) - Q'(v_i))_+$. For label-free[4] approximately truthful* strategies σ^K, we have for each agent i that

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

As $K \to \infty$, the approximations P_i^K converge to P_i, and the expectation goes to zero by Glivenko–Cantelli (since $\#U_i$ is finite). Then (4) follows by applying a union bound over the agents $i = 1, \ldots, n$.

To complete the proof, we simply check that permutation truthfulness implies approximate truthfulness*.[5] The key is to find a sufficiently large subset S of decision problems on which to apply the definition of permutation-truthfulness. This is guaranteed by the following combinatorial lemma.

Lemma 1. *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 $\pi\colon S \to S$ such that*
(i) $\hat{u}_i^k = u_i^{\pi(k)}$ *for all k in S;*
(ii) $\#S \ge K - Kd(\operatorname{marg} u_i, \operatorname{marg} \hat{u}_i) \cdot (\#U_i - 1)$.

Proof. Given u_i and $\hat{u}_i$ in U_i, construct a directed multigraph as follows. The vertex set 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 out-degree to a node with net in-degree; update the degrees of the new graph; and repeat until the graph is balanced, i.e., $\deg^+(v_i) = \deg^-(v_i)$ for all v_i in U_i. Let K' be the number

[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).

[5] In the special case where $\#U_i = 2$, every permutation-truthful strategy for agent i is necessarily approximately truthful. Thus, JS's Theorem 1 is correct with the original definition of approximate truthfulness if $|U_i| \le 2$ for every agent i. This is why we take $\#U_i = 3$ in our counterexample.

of new edges added by the this procedure. We have

$$\begin{aligned} K' &= \sum_{v_i} (\deg^+(v_i) - \deg^-(v_i))_+ \\ &= K \sum_{v_i} (\operatorname{marg}(v_i|u_i) - \operatorname{marg}(v_i|\hat{u}_i))_+ \\ &= Kd(\operatorname{marg} u_i, \operatorname{marg} \hat{u}_i). \end{aligned}$$

Now we have a balanced graph with $K + K'$ edges. Partition this graph into edge-disjoint cycles.[6] 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

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

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

References

Jackson, M. O. and H. F. Sonnenschein (2007): "Overcoming Incentive Constraints by Linking Decisions," *Econometrica*, 75, 241–257. [1]

[6]To do so, start at an arbitrary node, and form a path by arbitrarily selecting outgoing edges until the path contains a cycle. Remove the cycle and continue from the node where the cycle was completed. Since the graph is balanced, this process can terminate only when every edge has been removed.