האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

GIBBARD-SATTERTHWAITE SUCCESS STORIES AND OBVIOUS STRATEGYPROOFNESS

By

SOPHIE BADE AND YANNAI A. GONCZAROWSKI

Discussion Paper # 704 (October 2016)

מרכז פדרמן לחקר הרציונליות

THE FEDERMANN CENTER FOR THE STUDY OF RATIONALITY

Feldman Building, Edmond J. Safra Campus,
Jerusalem 91904, Israel
PHONE: [972]-2-6584135 FAX: [972]-2-6513681
E-MAIL: ratio@math.huji.ac.il
URL: http://www.ratio.huji.ac.il/

Gibbard-Satterthwaite Success Stories and Obvious Strategyproofness

Sophie Bade* Yannai A. Gonczarowski†

October 16, 2016

Abstract

The Gibbard-Satterthwaite Impossibility Theorem (Gibbard, 1973; Satterthwaite, 1975) holds that dictatorship is the only unanimous and strategyproof social choice function on the full domain of preferences. Much of the work in mechanism design aims at getting around this impossibility theorem. Three grand success stories stand out. On the domains of single peaked preferences, house matching, and of quasilinear preferences, there are appealing unanimous and strategyproof social choice functions. We investigate whether these success stories are robust to strengthening strategyproofness to obvious strategyproofness, recently introduced by Li (2015). A social choice function is obviously strategyproof implementable (OSP) implementable if even cognitively limited agents can recognize their strategies as weakly dominant.

For single-peaked preferences, we characterize the class of OSP-implementable and unanimous social choice rules as *dictatorships with safeguards against extremism* — mechanisms (which turn out to also be Pareto optimal) in which the dictator can choose the outcome, but other agents may prevent the dictator from choosing an outcome which is too extreme. Median voting is consequently not OSP-implementable. Indeed the only OSP-implementable quantile rules either choose the minimal or the maximal ideal point. For house matching, we characterize the class of OSP-implementable and Pareto optimal matching rules as *sequential barter with lurkers* — a significant generalization over bossy variants of bipolar serially dictatorial rules. While Li (2015) shows that second-price auctions are OSP-implementable when only one good is sold, we show that this positive result does not extend to the case of multiple goods. Even when all agents' preferences over goods are quasilinear and additive, no welfare-maximizing auction where losers pay nothing is OSP-implementable when more than one good is sold. Our analysis makes use of a gradual revelation principle, an analog of the (direct) revelation principle for OSP mechanisms that we present and prove.

*Department of Economics, Royal Holloway University of London, UK. *E-mail*: sophie.bade@rhul.ac.uk.

†Einstein Institute of Mathematics, Rachel & Selim Benin School of Computer Science & Engineering, and Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem, Israel; and Microsoft Research, *E-mail*: yannai@gonch.name.

1 Definitions

1.1 The Design Problem

There is a finite set of agents $N\colon = \{1, \ldots, n\}$ with typical element $i \in N$ and a set of outcomes Y. Agent i's preference R_i is drawn from a set of possible preference $\mathcal{R}_i$. Each possible preference R_i is a complete and transitive order on Y where xP_iy denotes the case that xR_iy but not yR_ix holds. If xR_iy and yR_ix, then x and y are R_i-indifferent. The domain of all agents' preferences is $\mathcal{R}\colon = \mathcal{R}_1 \times \cdots \times \mathcal{R}_n$. If two alternatives x and y are R_i-indifferent for each $R_i \in \mathcal{R}_i$, then x, y are **completely *i*-indifferent**. The set of all outcomes that are completely i-indifferent to y is $[y]_i$.

We consider three classes design problems. In case of quasilinear preferences the outcome space can be described as $Y\colon = X \times M$, where X is a set of allocations and M is a set of monetary payments with m_i representing the payment charged from agent i. Each agent's preference can be represented by a utility function $U_i(x, m) = u_i(x) + m_i$ where u_i is a utility function on X. In a political problem with **single peaked preferences**, we can represent the set of social choices as a the set of integers $\mathbb{Z}$. Moreover, for every $i \in N$ and any $R_i \in \mathcal{R}_i$, there exists an **ideal point** $y^* \in Y$ such that $y' < y \leq y^*$ or $y^* \geq y > y'$ implies yP_iy' for all $y, y' \in Y$. In a house matching problem, Y represents the set of matchings, constructed as follows. There exists a set of houses O, a house agent pair (o, i) is a match, a matching is a set of matches, where no house $o \in O$ or agent $i \in N$ partakes part in multiple matches. Given that we do not impose any requirement on the number of matches in a matching, this set Y is the set of matchings with outside options. The domain of preferences $\mathcal{R}$ for a house matching problem is such that each agent only cares about the house he is matched with. Each agent moreover strictly ranks any two houses. So x and y are completely i-indifferent if and only if i is matched to the same house under x and y. If y and y' are not completely i-indifferent then i must be matched with two different houses under x and y which he must therefore rank strictly.

A social choice function $scf : \mathcal{R} \to Y$ maps each profile of preferences $R \in \mathcal{R}$ to an outcome $scf(R) \in Y$.

1.2 Mechanisms

A (deterministic) **mechanism** is an extensive game form with the set N as the set of players. The set of **histories** H of this game form are the set of (finite and infinite[1]) paths from the root of the directed game form tree. For a history $h = (a^k)_{k=1,\ldots}$ of length at least L, we denote by $h|_L = (a^k)_{k=1,\ldots,L} \in H$ the **subhistory** of h of length $L \geq 0$. We write $h' \subset h$ when h' is a strict subhistory of h. A history is **terminal** if it is not a subhistory of any other history. (So the terminal histories correspond to paths to leaves and to infinite paths.)

[1]Unlike Li (2015), we allow for infinite histories mainly to allow for easier exposition of our analysis of the domain of single peaked preferences. See Section 6 for more details.

The set of all terminal histories is Z.

The set of possible actions after the nonterminal history h is $A(h) := \{a \mid (h,a) \in H\}$. The player function P maps any nonterminal history $h \in H \setminus Z$ to a player $P(h)$ who gets to choose from all actions $A(h)$ at h. Each terminal history $h \in Z$ is mapped to an outcome in Y.

Each player i has an **information partition** $\mathcal{I}_i$ of the set $P^{-1}(i)$ of all nodes h with $P(h) = i$, with $A(h) = A(h')$ if h and h' belong to the same cell of $\mathcal{I}_i$. The cell to which h with $P(h) = i$ belongs is $I_i(h)$.[2] A **behavior**[3] S_i for player i is an $\mathcal{I}_i$-measurable function mapping each h with $P(h) = i$ to an action in $A(h)$. A **behavior profile** $S = (S_i)_{i \in N}$ lists a behavior for each player. The set of behaviors for player i and the set of behavior profiles are respectively denoted $\mathcal{S}_i$ and $\mathcal{S}$. A behavior profile S induces a unique terminal history $h^S = (a^k)_{k=1,\ldots}$ s.t. $a^{k+1} = S_{P(h|_k)}(h|_k)$ for every k s.t. $h|_k$ is not terminal. The **mechanism** $M : \mathcal{S} \to Y$ maps the behavior profile $S \in \mathcal{S}$ to the outcome $y \in Y$ that is associated with the terminal history h^S. We call the set of all subhistories of the terminal history h^S the **path** $Path(S)$. A strategy $\mathbf{S}_i$ for agent i is a function $\mathbf{S}_i : \mathcal{R}_i \to \mathcal{S}_i$. The strategy profile $\mathbf{S}_i$ induces the social choice function $scf : \mathcal{R} \to Y$ if $scf(R)$ equals $M(\mathbf{S}(R))$ for each $R \in \mathcal{R}$.

In a **direct revelation mechanism** all agents move simultaneously. Agent i 's strategy space consists of his set of possible preferences $\mathcal{R}_i$.

1.3 Normative Criteria

A direct revelation mechanism $\hat{M}$ is **Pareto optimal** if it maps any R to an outcome $\hat{M}(R)$ that is Pareto optimal w.r.t. R. An outcome $y \in Y$ in turn is Pareto optimal w.r.t. R if there exists no $y' \in Y$ such that $y' R_i y$ holds for all i and $y' P_{i'} y$ holds for at least one i'.

A behavior S_i for agent i with preferences R_i is **dominant** if $M(S) R_i M(S'_i, S_{-i})$ holds for all alternative behaviors $S'_i \in \mathcal{S}_i$ for agent i and all possible behavior profiles S_{-i} for the other agents. A strategy $\mathcal{S}_i$ is dominant if it maps each R_i to a dominant behavior $\mathcal{S}_i(R_i)$. A direct revelation mechanism is incentive compatible if truthfully revealing one's preferences is a dominant strategy for each player. The revelation principle states that each social choice function that can be implemented in dominant strategies can be implemented by an incentive compatible direct revelation mechanism.

A behavior S_i is **obviously dominant** for agent i with preferences R_i if for every behavior profile S_{-i} for the other agents and for every $h \in Path(S)$ with $P(h) = i$, we have

$$M(S) R_i M(S') \text{ for all } S' \in \mathcal{S} \text{ with } h' \in Path(S'),\ I_i(h') = I_i(h), \text{ and } S'_i(h') \neq S_i(h).$$

If each behavior $\mathbf{S}_i(R_i)$ is obviously dominant for agent i with preference R_i then the strategy $\mathbf{S}_i$ is obviously dominant. The social choice function scf is implementable in obvi-

[2] Li (2015) also imposes the condition of perfect recall onto information partitions. Our results hold with and without prefect recall. For ease of exposition, we therefore do not impose prefect recall.

[3] We use the word behavior here since strategies are below defined as mapping from preferences to behaviors.

ously dominant strategies if $\mathbf{S}$ is a profile of obviously dominant strategies and if $scf(R) = M(\mathbf{S}(R))$ holds for all $R \in \mathcal{R}$. The revelation principle only applies in modified form to the case of implementation in obviously dominant strategies.

2 A Revelation Principle for Extensive-Form Mechanisms

A **gradual revelation mechanism** is a mechanism with the following additional features:

- Each cell $I_i(h)$ of each information partition $\mathcal{I}_i$ is a singleton.
- No agent has two directly consecutive choices: $P(h) \neq P(h, a)$ holds for every nonterminal history (h, a).
- Choices are real: no $A(h)$ is a singleton.
- Each nonterminal h and each $a \in A(h)$ are identified with sets $\mathcal{R}_i(h)$ and $\mathcal{R}_{P(h)}(a)$ for all $i \in N$. We let $\mathcal{R}_i(\emptyset) = \mathcal{R}_i$. For each h with $P(h) = i$, the set $\{\mathcal{R}_i(a) \mid a \in A(h)\}$ partitions $\mathcal{R}_i(h)$ and let $\mathcal{R}_i(h, a) \colon = \mathcal{R}_i(a)$. If $P(h) \neq i$ then let $\mathcal{R}_i(h) = \mathcal{R}_i(h, a)$ for all $a \in A(h)$.

  Finally, if $h \neq \emptyset$ is a nonterminal history then $\mathcal{R}_i(h) = \mathcal{R}_i(a)$, where a is the last action taken by i in h.
- For every nonterminal h, if $\{M(S_i, S_{-i}) \mid h \in Path(S_i, S_{-i})\}$ is a nonempty set of completely i-indifferent outcomes wit $P(h) = i$, then $\mathcal{R}(S_i(h))$ is a singleton.

A strategy $\mathbf{T}_i$ for player i in a gradual revelation mechanism is a **truthtelling strategy** if $R_i \in \mathcal{R}_i(\mathbf{T}_i(R_i)(h))$ holds for all R_i and nonterminal h with $P(h) = i$ and $R_i \in \mathcal{R}_i(h)$. So $\mathbf{T}_i$ is a truthtelling strategy if agent i reveals which set his preference R_i belongs to, whenever possible. If $R_i \notin \mathcal{R}_i(h)$, then the definition imposes not restriction on the behavior of agent $i = P(h)$ with preference R_i. Since the specification of $\mathbf{T}_i(R_i)$ for histories h with $R_i \notin \mathcal{R}_i(\mathbf{T}_i(R_i)(h))$ is inconsequential to our analysis, we call any truthtelling strategy *the* truthtelling strategy. A gradual revelation mechanism is **obviously incentive compatible** if the truthtelling strategy $\mathbf{T}_i$ is obviously dominant for each agent i. We say that an obviously incentive compatible gradual revelation mechanism M implements a social choice function $scf : \mathcal{R} \to Y$ if $scf(\cdot) = M(\mathbf{T}(\cdot))$.

Theorem 1 *A social choice function is implementable in obviously dominant strategies if and only if some obviously incentive compatible gradual revelation mechanism M implements it.*

Proof Fix any mechanism $M : \mathcal{S} \to Y$, social choice function $scf : \mathcal{R} \to Y$ and obviously strategyproof strategy profile $\mathbf{S} : \mathcal{R} \to \mathcal{S}$. Define a new mechanism $M^1 : \mathcal{S}^1 \to Y$ that is identical to M except that all information sets are singletons. In M, a strategy for agent i is an $\mathcal{I}_i$-measurable function mapping a nonterminal history $h \in P^{-1}(i)$ of player i to the actions $A(h)$ available at h. In M^1, a strategy for agent i has the same definition, only without the requirement of being $\mathcal{I}_i$-measurable. So, we have $\mathcal{S} \subset \mathcal{S}^1$. Since $\mathbf{S}$ is obviously strategyproof in M, we obtain that $\mathbf{S}$ is also obviously strategyproof in M^1 (as the relations that must hold for the latter are a subset of those that must hold for the former).

For every nonterminal history h in M^1, let $i = P(h)$, set $\mathcal{R}_i(h) = \{R_i \in \mathcal{R}_i \mid \exists S_{-i} : h \in Path(S^1(R_i), S_{-i})\}$, and for every $a \in A(h)$, set $\mathcal{R}_i(a) = \{R_i \in \mathcal{R}_i(h) \mid \mathbf{S}_i(R_i)(h) = a\}$. We note that $\{\mathcal{R}_i(a) \mid a \in A(h)\}$ partitions $\mathcal{R}_i(h)$, however some of the sets in this partition may be empty. Since $\mathbf{S}$ is obviously strategyproof in M^1, $\mathbf{T}$ is obviously strategyproof in M^1 as well (w.r.t. the maps just defined). Furthermore, $M^1(\mathbf{T}(\cdot)) = M^1(\mathbf{S}(\cdot)) = scf(\cdot)$.

In the following steps of the proof, we describe modifications to the game tree of the mechanism. For ease of presentation, we consider the maps P, $S_i(R_i)$ and $\mathcal{R}_i$ to be defined over nodes of the tree (rather than histories, as we may modify the paths to these nodes).

Let $M_0^2 = M^1$. For every agent $i \in N$ (inductively on i), we define a new mechanism $M_i^2 : \mathcal{S}_i^2 \to Y$ as follows: For every preference R_i, for every minimal history h s.t. $P(h) = i$ and $\{M_{i-1}^2(\mathbf{T}(R)) \mid h \in Path(\mathbf{T}(R))\}$ is a nonempty set of completely i-indifferent outcomes, let $a = \mathbf{T}_i(R_i)(h)$, remove R_i from the set $\mathcal{R}_i(a)$, and put $\mathcal{R}_i(a') = \{R_i\}$ for a new action a' at h that leads to a subtree that is a duplicate of that to which a leads before this change (with all maps from the nodes of the duplicate subtree defined as on the original subtree). Note that $M_i^2(\mathbf{T}(R)) = M_{i-1}^2(\mathbf{T}(R))$ holds for all R, so we have $M_i^2(\mathbf{T}(\cdot)) = M_{i-1}^2(\mathbf{T}(\cdot)) = scf(\cdot)$. Moreover, since $\mathbf{T}$ is obviously strategyproof in M_{i-1}^2, $\mathbf{T}$ is obviously strategyproof in M_i^2. Set $M^2 = M_n^2$.

Define a new mechanism $M^3 : \mathcal{S}^3 \to Y$ by dropping from M^2 any action a for which there exists no R such that a is on the path $Path(\mathbf{T}(R))$ in M^2. Since $\mathbf{T}$ is obviously strategyproof in M^2, $\mathbf{T}$ is also obviously strategyproof in M^3. Furthermore, $M^3(\mathbf{T}(\cdot)) = M^2(\mathbf{T}(\cdot)) = scf(\cdot)$.

Define a new mechanism $M^4 : \mathcal{S}^4 \to Y$ as follows. Identify a maximal set of histories H^* in M^3 that satisfies all of the following:

- Each $h \in H^*$ is either nonterminal or infinite.
- $P(h) = i$ for all nonterminal $h \in H^*$ and some i,
- there exists a history $h^* \in H^*$ such that $h^* \subset h$ for all $h \in H$, and
- if $h \in H^*$ then $h' \in H^*$ for all h' with $h^* \subset h' \subset h$.

"Condense" each such H^* by replacing the set of actions $A(h^*)$ at h^* s.t. at h^*, agent i directly chooses among all possible nodes (h, a), where h is a maximal nonterminal history in H^* and $a \in A(h)$; in addition, for every infinite history h in H^*, add an action to $A(h^*)$

that chooses a new leaf with the same outcome as h. For every new action a' that chooses a node (h, a) from M^3, set $\mathcal{R}_i(a') = \mathcal{R}_i(a)$; for every new action a' that chooses a new leaf with the same outcome as in an infinite history $h = (a^k)_{k=1,\dots}$ of M^3, set $\mathcal{R}_i(a') = \cap_k \mathcal{R}_i(a^k)$. Since $\mathbf{T}$ is obviously strategyproof in M^3, $\mathbf{T}$ is also obviously strategyproof in M^4. Furthermore, $M^4(\mathbf{T}(\cdot)) = M^3(\mathbf{T}(\cdot)) = scf(\cdot)$.

Define a new mechanism $M^5 : \mathcal{S}^5 \to Y$ as follows. Identify a maximal set of histories H^* in M^4 that satisfies all of the following:

- $|A(h)| = 1$ for all nonterminal $h \in H^*$,
- there exists a history $h^* \in H^*$ such that $h^* \subset h$ for all $h \in H$, and
- if $h \in H^*$ then $h' \in H^*$ for all h' with $h^* \subset h' \subset h$.

"Condense" each such H^* by replacing the subtree rooted at the node h^* with the subtree rooted at the node h, where h is the maximal history in H^*. If the h is infinite, then replace h^* with a new leaf with the same outcome as h. Since $\mathbf{T}$ is obviously strategyproof in M^4, $\mathbf{T}$ is also obviously strategyproof in M^5. Furthermore, $M^5(\mathbf{T}(\cdot)) = M^4(\mathbf{T}(\cdot)) = scf(\cdot)$.

By construction, M^5 is a gradual revelation mechanism that implements scf. $\square$

2.1 Preliminary Analysis

For any nonterminal history h, let $Y(h)$ be the set of all outcomes associated with a terminal history h' with $h \subset h'$.

In an obviously incentive compatible gradual revelation mechanism M, let h be a nonterminal history and let $i = P(h)$. We define $Y_h^* \subseteq Y(h)$ to be the set of outcomes y such that there exists some $a \in A(h)$ s.t. $Y(h, a) \subset [y]_i$.[4] We define $A_h^* \subseteq A(h)$ to be the set of actions a such that $Y(h, a) \subset [y]_i$ for some $y \in Y_h^*$. We call A_h^* the set of **dictatorial** actions at h. Let $\overline{A_h^*} \triangleq A(h) \setminus A_h^*$ and $\overline{Y_h^*} \triangleq Y(h) \setminus Y_h^*$. We call $\overline{A_h^*}$ the set of **nondictatorial** actions at h. We will show below that Y_h^* and A_h^* are non-empty for the single peaked as well as the matching domain. (See Lemmas 3.2, 6.3, 6.7 and 6.8, and Theorem 4.) Before considering these domains, we first prove a general lemma.

Lemma 2.1 *Fix an obviously incentive compatible gradual revelation mechanism M. Let h be a nonterminal history and let $i = P(h)$. If there exists $y \in Y(h)$ s.t. $[y]_i \cap Y(h, a) \neq \emptyset \neq [y]_i \cap Y(h, a')$ for two distinct $a, a' \in A(h)$, and furthermore there exists $R_i \in \mathcal{R}_i(h)$ s.t. R_i ranks $[y]_i$ at the top among $Y(h)$, then $[y]_i \cap Y(h) \subset Y_h^*$.*

Proof Suppose not, and assume w.l.o.g. that $y \notin Y_h^*$. Since $y \notin Y_h^*$, there exists a preference profile R_{-i} with $h \in Path(\mathbf{T}(R))$ such that $M(\mathbf{T}(R)) = y' \notin [y]_i$. Assume w.l.o.g.

[4] To reduce clutter, by slight abuse of notation we here and henceforth write $Y(h, a)$, and sometimes $Y(a)$, instead of $Y\big((h, a)\big)$.

that $\mathbf{T}_i(R_i)(h) \neq a'$. Since $[y]_i \cap Y(h, a') \neq \emptyset$, there exists a preference profile R' with $h \in Path(\mathbf{T}(R'))$ and $\mathbf{T}(R')(h) = a'$ such that $M(\mathbf{T}(R')) \in [y]_i$. Since $M(\mathbf{T}(R')) R_i y P_i y' = M(\mathbf{T}(R))$, the mechanism is not obviously strategyproof, reaching a contradiction. □

3 House Matching

In the house-matching problem $(Y, \mathcal{R})$ derived from sets of houses O, the possible outcomes are one-to-one matchings between the agents and houses in a set O. We assume for simplicity that there are at least as many houses as there are agents, and that all agents must be matched. (See Section 3.1 for a discussion on what happens if this is not the case). The preferences of each agent over outcomes is only determined according to the house that this agent is matched with. Under this setting, we show that a Pareto optimal social choice function is OSP-implementable if and only if it is implementable via **sequential barter with lurkers**. Before we define sequential barter with lurkers, we first review some prior work and take a look at a special case, **sequential barter**.

Under the setting described above, Li (2015) shows that the popular (and Pareto optimal) top trading cycles (TTC) mechanism (Shapley and Scarf, 1974) is not obviously strategyproof, yet that serial dictatorship is. Under certain conditions, Pycia (2016) shows that curated dictatorship, a bossy variant of serial dictatorship, is the unique OSP-implementable rule. Ashlagi and Gonczarowski (2016) show that all Pareto optimal stable matching rules are OSP implementable via the following generalization of bipolar serially dictatorial rules (Bogomolnaia et al., 2005): At each mechanism step, either choose an agent and give her free choice among all unmatched houses, or choose two agents, partition all unmatched houses into two sets, and each of the agents gets priority in one of the sets, i.e., gets free pick from that set. If any agent chose from her set, then the other gets to pick from all remaining houses. If both agents did not choose from their sets, then each gets her favorite choice (which is in the set of the other). Troyan (2016) generalizes even further by showing that TTC where at any given point in time no more than two agents are pointed to, is OSP implementable, and shows that no other TTC mechanism is OSP implementable. Before presenting our characterization for OSP-implementable Pareto optimal social choice rules, we first present the OSP-implementable and Pareto optimal **sequential barter** mechanism that generalizes all of the above, but still does not capture the full gamut of OSP-implementable and Pareto optimal social choice rules (that are captured by sequential barter with lurkers — a significant extension defined below). In sequential barter, somewhat similarly to TTC, each agent points to her preferred house, however the houses point to agents gradually: at each turn, the mechanism chooses (based only on the preferences of already-matched agents) an item and that item points to some agent, under the restriction that at most two agents are pointed at. Once a cycle forms, the agents and houses in that cycle are matched, and the process continues. (As noted above, this is a richer set of mechanisms than TTC with at most two agents pointed to at any given time, as houses can defer pointing to an agent and then choose that agent based on preferences of agents that have already been matched, or even choose that agent based nonbossily only on matched pairs already formed). With an eye toward our full characterization, we now present a different characterization of sequential barter, which will help us present the full characterization, i.e., sequential barter with lurkers.

A mechanism is a sequential barter mechanism iff it is equivalent to a mechanism of the following form:

Sequential Barter

1. Initialization:
   (a) O is initialized to be the set of all houses (O tracks the set of unmatched houses).
   (b) $T \leftarrow \emptyset$ (this is the set active traders — see below).[5]
2. Mechanism progress: as long as there are unmatched agents, perform an endowment step (defined below).
3. Endowment step:
   (a) Choose an unmatched agent i (all choices made by the mechanism are functions of preferences already disclosed to the mechanism), subject to the following constraint:
      - If $|T| = 2$, then $i \in T$.

   (b) If $i \notin T$, then update $T \leftarrow T \cup \{i\}$ and initialize $D_i \leftarrow \emptyset$.
   (D_i is the set of houses i can claim — see below.[6])
   (c) Choose some $\emptyset \neq H \subseteq O_i \setminus D_i$.
   (d) Update $D_i \leftarrow D_i \cup H$ and perform a question step (defined below) for i.
4. Question step for an agent i:
   (a) Ask i whether the house she prefers most among O_i is in D_i. If so, then perform a matching step (defined below) for i and that house.
5. Matching step for an agent i and a house o:
   (a) Match i and o.
   (b) Update $T \leftarrow T \setminus \{i\}$ and $O \leftarrow O \setminus \{o\}$.
   (c) i discloses her full preferences to the mechanism (and leaves the game).

[5]An invariant of the mechanism is that $|T| \leq 2$.
[6]An invariant of the mechanism is that $D_j \subseteq O$ for every $j \in T$.

(d) If $T \neq \emptyset$, then for the unique agent $j \in T$:

i. If $o \in D_j$, then set $D_j \leftarrow O_j$.

ii. Perform a question step for j.

As it turns out, despite all previously-known OSP and Pareto optimal mechanisms for house matching being special cases the of above mechanism, the above mechanism is in fact not the only form of possible OSP and Pareto optimal mechanisms. Indeed, Fig. 1 presents an OSP and PO mechanism where three traders are active at some history. (We note that

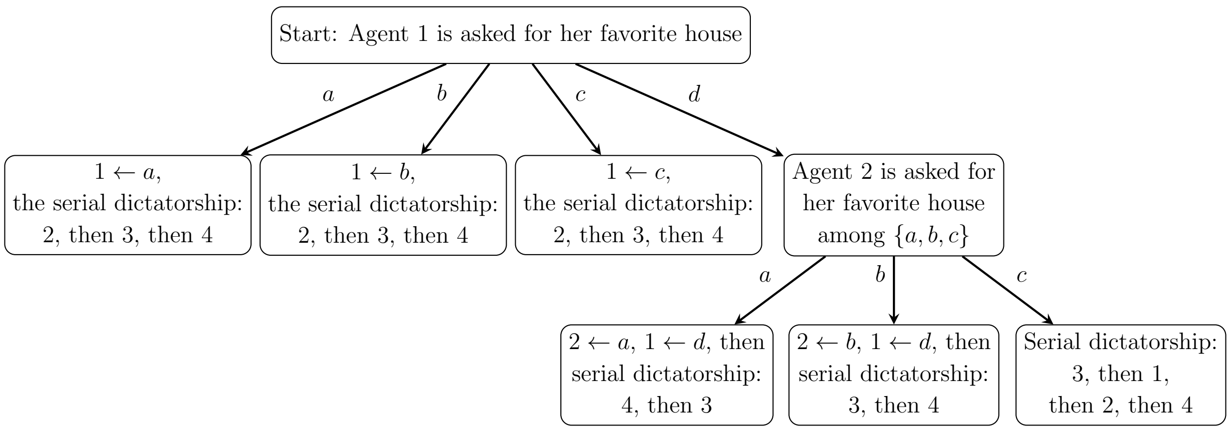

Figure 1: An OSP and Pareto optimal mechanism for four agents $1, 2, 3, 4$ and four houses a, b, c, d, with three active agents when agent 3 chooses at the bottom-right.

this mechanism is bossy.) The crux of the mechanism in Fig. 1 is that since the mechanism offers agent 1 free choice of houses a, b, and c, if agent 1 chooses not to claim any of them, then the mechanism infers that this agent prefers house d most. Therefore, the mechanism can safely decide that agent 2 never gets house d without violating Pareto optimality (the competition over house d is now between agent 1 and 3). So, if the mechanism knows that some agent prefers some house the most (the only OSP way to do so without offering this house to this agent is to offer all other possible houses to this agent; in this case, we say that this agent is a *lurker* for that house), then the mechanism may decide not to allow some other agents to ever get this house, and this allows the introduction of additional traders (beyond two traders) under certain delicate constraints. We are now ready to present our characterization of OSP-implementable and Pareto optimal social choice rules. A mechanism of sequential barter with lurkers is of the following form:

Sequential Barter with Lurkers

1. Initialization:

(a) O is initialized to be the set of all houses (O tracks the set of unmatched houses).

(b) $T \leftarrow \emptyset$ (this is the set active traders — see below).[7]

[7] An invariant of the mechanism is that $|T \setminus L| \leq 2$.

(c) $L \leftarrow \emptyset$ (this is the set of lurkers, i.e., active traders who lurk houses — see below).

(d) $G \leftarrow O$ (this is the set of houses that have no lurkers — see below).[8]

2. Mechanism progress: as long as there are unmatched agents, then run a mechanism round as follows:

   (a) Perform an endowment step (defined below). Let i be the agent chosen in this endowment step.

   (b) Identify lurkers, in preparation for the next round:

      i. If i is still unmatched after the above endowment step (including all steps directly or indirectly triggered by this question step), then perform a lurker check step (defined below) for i.

      ii. If $i \notin T \setminus L$,[9] and if $T \setminus L \neq \emptyset$, then perform a lurker check step for the unique agent $j \in T \setminus L$.[10]

3. Endowment step:

   (a) Choose an unmatched agent i (all choices made by the mechanism are functions of preferences already disclosed to the mechanism), subject to the following constraint:

      - If $|T \setminus L| = 2$, then $i \in T$.

   (b) If $i \notin T$, then update $T \leftarrow T \cup \{i\}$ and initialize D_i and O_i as follows:

      i. Choose either $O_i \leftarrow G$ or $O_i \leftarrow O$, subject to the following constraint:

         - If $T \setminus L = \{i, j\}$ for some agent j, then it is not the case that both $O_i \neq G$ and $O_j \neq G$.

         (O_i is the set of houses that may be possible for i.[11])

      ii. Initialize $D_i \leftarrow \emptyset$ (this is the set of houses i can claim — see below).[12]

   (c) Choose some $\emptyset \neq H \subseteq O_i \setminus D_i$ subject to the following constraints: (If it is not possible to choose such nonempty H subject to these constraints, then stop this endowment step, start a new endowment step, and choose a different agent there.)

      - If $H \setminus G \neq \emptyset$, then $H = O_i \setminus D_i$.
      - If $O_t \neq G$ for $t \in T \setminus (L \cup \{i\})$, then $H \cap D_t = \emptyset$.

   (d) Update $D_i \leftarrow D_i \cup H$ and perform a question step (defined below) for i.

[8] An invariant of the mechanism is that $G = O$ iff $L = \emptyset$.

[9] So i became a lurker or matched or during this endowment step.

[10] In this case, the lurker check step cannot result in new matches because j is the only agent in $T \setminus L$.

[11] An invariant of the mechanism is that for every $j \in T \setminus L$, either $O_j = G$ or $O_j = O$, and if these options differ (i.e., if $L \neq \emptyset$), then the latter is possible for at most one agent $j \in T \setminus L$.

[12] An invariant of the mechanism is that $D_j \subseteq O_j$ for every $j \in T$.

4. Question step for an agent i:

   (a) Ask i whether the house she prefers most among O_i is in D_i. If so, then perform a matching step (defined below) for i and that house.

5. Matching step for an agent i and a house o:

   (a) Match i and o.

   (b) Update $T \leftarrow T \setminus \{i\}$ and $O \leftarrow O \setminus \{o\}$, and $O_j \leftarrow O_j \setminus \{o\}$ for every $j \in T$.

   (c) If $i \in L$, then update $L \leftarrow L \setminus \{i\}$.

   (d) If $o \in G$, then update $G \leftarrow G \setminus \{o\}$.

   (e) i discloses her full preferences to the mechanism (and leaves the game).

   (f) For every agent $j \in T$:[13]

      i. If $o \in D_j$, then set $D_j \leftarrow O_j$.

      ii. Perform a question step for j.

6. Lurker check step for an agent i:

   (a) If $O_i = G$ and $O_i \setminus D_i = \{o\}$ for some house o, then[14] i becomes a *lurker* for o:

      i. Update $L \leftarrow L \cup \{i\}$ and $G \leftarrow G \setminus \{o\}$.

      ii. If $|L \setminus T| = 2$ then let j be the unique agent $j \in T \setminus (L \cup \{i\})$ and choose to either keep O_j as is or to update update $O_j \leftarrow O_j \setminus \{o\}$, subject to the following constraints:

         - After updating, it must be that $O_j = O$ or $O_j = G$.
         - After updating, it must be that $D_j \subseteq O_j$.

      iii. If O_j was updated, then perform a question step for j.

Theorem 2 *A Pareto-optimal social choice function in a house-matching problem is OSP-implementable if and only if it is implementable via sequential barter with lurkers.*

The proof of Theorem 2 follows from Lemmas 3.1 through 3.12.

Lemma 3.1 *Any social choice function implemented via sequential barter with lurkers is Pareto-optimal. Moreover, such a social choice function can be OSP-implemented.*

Proof Pareto optimality follows from the fact that whenever a set of houses leave the game, then one of them is most-preferred by its matched agent among all not-previously-matched houses, another is most-preferred by its matched agent among all not-previously-matched houses except for the first, etc.

[13]The outcome of the mechanism does not depend on the order of traversal of $T \cup L$. This insight is what ensures that the mechanism is OSP.

[14]Note that in this case, i prefers o most among O_i.

The OSP implementation follows the above very closely. Ashlagi and Gonczarowski (2016) have already used in their examples that asking an agent whether he most prefers some given house, and if the agent's answer is "yes", assigning to her that house (and otherwise continuing), is OSP as long as the agent is assured they will always get "at least" that house. The only thing that has to be checked is that the above "rules" (promising agents i to be able to claim their top choice from D_i after O_i is reduced, and to claim any house from O_i if a house from D_i becomes matched) can indeed be met simultaneously (see Footnote 13) for all agents. This holds because of the reduction of O_j for all nonlurkers j but at most one whenever an agent becomes a lurker, and by the restrictions on the choice of H. □

Theorem 1 shows that any OSP implementable rule is implementable via an obviously incentive compatible gradual revelation mechanism. So, to show the converse direction, it is sufficient to show that any Pareto optimal rule that is implementable via an obviously incentive compatible gradual revelation mechanism corresponds to sequential barter with lurkers. For the reminder of the proof, let M be an obviously incentive compatible gradual revelation mechanism that implements a Pareto optimal rule. Since $\mathcal{R}$ is finite, all histories in M are finite.

Definition 3.1 *Let h be a history of M and $i \in N$ an agent.*

1. *$O_i(h)$ is the set of houses that are assigned to i in any terminal history with h as a subhistory.*

2. *$D_i(h)$ is the set of houses o that agent i can assure at some subhistory $h' \subset h$, so we have $P(h') = i$ and $o \in Y^*_{h'}$.*[15]

3. *$D^<_i(h)$ is the set of houses o that agent i can assure at some* strict *subhistory $h' \subsetneq h$, so we have $P(h') = i$ and $o \in Y^*_{h'}$.*

4. *Agent i moves at h if $P(h) = i$. Agent i moves before h if $P(h') = i$ holds for a strict subhistory h' of h.*

5. *Agent i is matched to house o at h if $O_i(h) = \{o\}$.*

6. *A preference R_i that ranks house o at the top is denoted $R_i : o$, similarly $R_i : o, o'$ denotes a preference that ranks o above o' which it ranks in turn above all other houses. The definition is extended to lists of any length. The preference $R_i : o, \ldots, o^*$ ranks o at the top and o^* at the bottom. Once again, this notation is extended to lists of any length.*

[15]By abuse of notation, in this section we define Y^*_h to be the set of *houses* o (rather than outcomes) such that there exists some $a \in A(h)$ s.t. $P(h)$ is matched with o in all outcomes in $Y(h, a)$. Due to the nature of preferences in the matching domain, this (simpler) projection of the general definition of Y^*_h suffices for the analysis of this section and helps simplify presentation.

Remark 1 *Note that $O_i(h)$ is (weakly) decreasing in h, while $D_i(h)$ and $D_i^<(h)$ are both (weakly) increasing in h. Note also that $D_i^<(h)$ is always a subset of $D_i(h)$, and that i equals $P(h)$ if $D_i(h)$ and $D_i^<(h)$ differ.*

Lemma 3.2 *Fix a history h of M and let $i = P(h)$.*

1. *There exists at most one action $a \in \overline{A_h^*}$ — denote it by $\tilde{a}$.*

2. *The truthtelling strategy at h is such that*

    - *For every $o \in Y_h^*$: $\bigcup_{a \in A(h):O_i(h,a)=\{o\}} \mathcal{R}_i(h,a) = \{R_i \in \mathcal{R}_i(h) \mid o = \max_{R_i} O_i(h)\}$,*
    - *$\mathcal{R}_i(h,\tilde{a}) = \mathcal{R}_i(h) \setminus \left(\bigcup_{a \in A(h)\setminus\{\tilde{a}\}} \mathcal{R}_i(h,a)\right)$.*

Proof of Lemma 3.2 We start with a preliminary observation, which states a sufficient condition for $\overline{A_h^*}$ to be a singleton.

(*) Fix h, i such that $P(h) = i$. If for all $o, o' \in O_i(h) \setminus D_i(h)$ with $o \neq o'$ there exists some $R_i \in \mathcal{R}_i(h)$ and with $R_i : o, o'$, then $\overline{A_h^*}$ is a singleton.

Suppose not, that is suppose there exist two different actions a and a' in $\overline{A_h^*}$. Since M is a gradual revelation mechanism, $\mathcal{R}_i(h,a) \neq \emptyset$. Say that $R_i \in \mathcal{R}_i(h,a)$ and let $o = \max_{R_i} O_i(h)$. Since $a \in \overline{A_h^*}$, there exist preferences R_{-i} for all agents other than i such that $(h,a) \in Path(M(\mathbf{T}(R)))$ and $M(\mathbf{T}(R))(i) = \tilde{o}$ for some $\tilde{o} \neq o$. Note that $o \notin D_i(h)$, for otherwise agent i with preferences R_i is better off to deviate from $\mathbf{T}_i(R_i)$ given that all other agents follow $\mathbf{T}_{-i}(R_{-i})$, a contradiction to the strategyproofness of M. By the same reasons there exists a $R_i' \in \mathcal{R}_i(h,a')$ with $o' = \max_{R_i'} O_i(h) \notin D_i(h)$. If $o = o'$, then Lemma 2.1 which requires $o' \notin O_i(h,a)$ is contradicted. If $o \neq o'$ then there is no obviously strategy proof strategy for $R_i'' : o, o'$, which, by assumption, is in $\mathcal{R}_i(h)$. To see this, observe that for any strategy choosing a at h agent i may be matched house in $O_i(h,a) \setminus \{o\}$, while for some strategy choosing a' at h agent i may (for some R_{-i}) be matched with o'. Since o' is strictly R_i''-preferred to any house in $O_i(h,a) \setminus \{o\}$ an obviously strategy proof strategy for R_i'' must prescribe an action other than a at h. However i is not matched to o for any such strategy, while he may be matched with o for some strategy that chooses a at h for some R_{-i}. There is, in sum no obviously incentive compatible strategy for i with preferences R_i.

We now fix an agent i and prove both parts of the lemma by full induction over histories h with $P(h) = i$. We first note that the condition in Observation (*) is satisfied at every minimal history h^* with $P(h^*) = i$, as $\mathcal{R}_i(h^*)$ equals the set of all possible preferences $\mathcal{R}_i$. (Therefore, Part 1 holds at h^*.) We next show that Part 2 holds at a history h if Part 1 does. Finally, we argue that the condition in Observation (*) holds at h if Parts 1 and 2 hold for all strict subhistories $h' \subsetneq h$ with $P(h') = i$.

To see that Part 2 holds at h with $P(h) = i$ whenever Part 1 does, note that $\mathcal{R}_i(h,a)$ is non-empty for all $a \in A(h)$ as M is a gradual revelation mechanism. If Part 1 holds at

h, we have by definition that $\mathcal{R}_i(h,\tilde{a}) = \mathcal{R}_i(h) \setminus \left(\bigcup_{a\in A(h)\setminus\{\tilde{a}\}} \mathcal{R}_i(h,a)\right)$. Fix $o \in Y_h^*$. To see that every $R_i \in \mathcal{R}_i(h)$ with $o = \max_{R_i} O_i(h)$ chooses an action a with $O_i(h,a) = \{o\}$, note that i obtains o (for sure) if choosing such an action a at h (and such an action exists, since $o \in Y_h^*$), whereas for any a' with $O_i(h,a') \neq \{o\}$, agent i either certainly or under some scenario obtains a house o' other than o (which is strictly R_i-worse than o). To see that an action a with $O_i(h,a) = \{o\}$ is chosen only by $R_i \in \mathcal{R}_i(h)$ with $o = \max_{R_i} O_i(h)$, let a be such an action and let R_i with $\max_{R_i} O_i(h) = o'$ for some $o' \neq o$. We will show that $\mathbf{T}_i(R_i)$ cannot prescribe to choose a at h. Indeed, since $o' \in O_i(h)$, there must exist some R' with $h \in \mathit{Path}(M(\mathbf{T}(R')))$ and $M(\mathbf{T}(R'))(i) = o'$. Let a' be the choice prescribed by $\mathbf{T}_i(R'_i)$ at h, and note that $a' \neq a$ since $o' \in O_i(h,a')$ but $o' \notin O_i(h,a)$. While agent i with preferences R_i obtains her most preferred feasible house o' under the best possible scenario if choosing a' at h, he does not obtain o' under the worst possible case scenario if he chooses a at h (as he obtains o when choosing a at h), and therefore she does not choose a, as required.

To see that Part 1 holds at h with $P(h) = i$ if Parts 1 and 2 holds for all strict subhistories $h' \subsetneq h$ with $P(h') = i$, we show that the sufficient condition identified in Observation (*) holds for every such h. The condition is trivially satisfied if $O_i(h) \setminus D_i(h)$ is a singleton. So suppose there exist two distinct $o, o' \in O_i(h) \setminus D_i(h)$; therefore, by Part 1 for the induction hypothesis, agent i chooses $\tilde{a}$ in all strict subhistories of h. Since $D_i(h)$ and $O_i(h)$ respectively are weakly increasing and decreasing, by Part 2 for the induction hypothesis $\mathcal{R}_i(h)$ contains at least all R_i that rank $o \in O_i(h) \setminus D_i(h)$ above all $D_i(h)$. So, $\mathcal{R}_i(h)$ in particular contains (all preferences of the form) $R_i : o, o'$ and the sufficient condition cited in Observation (*) holds. $\square$

Definition 3.2 *Let h be a history of M and let $i \in N$.*

1. *Agent i is* active *at h if $D_i(h) \neq \emptyset$ and $|O_i(h)| > 1$.*
2. *Agent i can force a house o at h if there exists a strategy of i that from h guarantees that i will be matched with o. (I.e., there exists a preference R_i such that $o = M(\mathbf{T}(R))(i)$ holds for all R_{-i} with $h \in \mathit{Path}(M(\mathbf{T}(R)))$).*
3. *Agent i is a* dictator *at h if he can force at h every house from $O_i(h)$. (I.e., if for every profile R s.t. $h \in \mathit{Path}(\mathbf{T}(R))$, it is the case that $M(\mathbf{T}(R))(i)$ depends only on R_i.)*

Remark 2 *Note that we do not assume that $i = P(h)$ in any of the parts of Definition 3.2.*

Lemma 3.3 *Fix a history h of M and an agent i who has moved before or at h.*

1. *If $o = \max_{R_i} O_i(h) \in D_i(h)$ holds for some $R_i \in \mathcal{R}_i(h)$, then i can force o at h.*
2. *If i is not a dictator at h then $D_i(h) \subsetneq O_i(h)$.*

3. *If $O_i(h) \subseteq D_i(h)$, then i is a dictator at h.*

Proof of Lemma 3.3

1. Since $o \in D_i(h)$ there is a history $h' \subset h$ with $P(h') = i$ and $o \in Y^*_{h'}$. Since i may deviate to a strategy that chooses an action a at h' with $O_i(h', a) = \{o\}$, agent i with preference R_i must weakly prefer $M(\mathbf{T}(R))(i)$ to o for any R_{-i} such that $h \in M(\mathbf{T}(R))$. Therefore, since $o = \max_{R_i} O_i(h)$ agent i must be able to force o at h.

2. Since i is not a dictator at h, there exists a preference $R_i \in \mathcal{R}_i(h)$ such that i cannot force at h for himself the house o ranked highest among $O_i(h)$ by R_i. By Part 1, $o \notin D_i(h)$. Since i has moved before or at h, we have that $D_i(h) \neq \emptyset$; let $d \in D_i(h)$. Consider the preference $R_i : o, d$. Since $o \in O_i(h) \setminus D_i(h)$, by Lemma 3.2 we have that $R_i \in \mathcal{R}_i(h)$. Since i cannot force the house o, there exist preferences R_{-i} for all agents but i such that $h \in Path(M(\mathbf{T}(R)))$ and $M(\mathbf{T}(R))(i) \neq o$. As in the proof of Part 1, since $d \in D_i(h)$, by strategyproofness $M(\mathbf{T}(R))(i)$ must be weakly R_i-preferred to d, implying $M(\mathbf{T}(R))(i) = d$ and therefore $d \in O_i(h)$.

3. Immediately by Part 1, as $\max_{R_i} O_i(h) \in D_i(h)$ holds for all $R_i \in \mathcal{R}_i(h)$

□

Definition 3.3 *Let h be a history of M.*

- *An agent i who has moved before h, but is not a dictator at h is a* lurker *for a house o at h if at the maximal* strict *subhistory $h' \subsetneq h$ with $P(h') = i$, it is the case that $o \in O_i(h')$ and $D_i(h') = O_i(h') \setminus \{o\}$.*

- *$O(h)$ is the union of the sets $O_i(h)$ for all agents i that are not yet matched at h.*

- *$G(h)$ is the set of houses in $O(h)$ that have no lurkers at h.*

Lemma 3.4 *Fix a history h in M and an agent i who is not yet matched at h. If agent i is not a lurker for some house o at h, then there exists $o' \neq o$ s.t. $R_i \in \mathcal{R}_i(h)$ for every $R_i : o'$.*

Proof of Lemma 3.4 Assume that i not a lurker for o. If $D_i(h) = \emptyset$, then since $|O_i(h)| > 1$ (since i is not yet matched), there exists $o' \in O_i(h) \setminus \{o\}$, and by definition for every $R_i : o'$, it is the case that $R_i \in \mathcal{R}_i = \mathcal{R}_i(h)$, where the equality is by Lemma 3.2 since $D_i(h) = \emptyset$. If i is a dictator at h, then since $|O_i(h)| > 1$, there is more than one possible top choice for the preferences of i among $O_i(h)$. Assume therefore that $D_i(h) \neq \emptyset$ and that i is not a dictator at h, and let $h' \subsetneq h$ be the maximal strict subhistory of h with $P(h') = i$. Since i is not a dictator at h, she is not a dictator at h', and so by Lemma 3.3(2), $D_i(h') \subsetneq O_i(h')$. Since i is not a lurker for o, we have that there exists $o' \in O_i(h') \setminus D_i(h')$ s.t. $o' \neq o$. By Lemma 3.2, we have that $R_i \in \mathcal{R}_i(h', \tilde{a}) = \mathcal{R}_i(h)$ for every $R_i : o'$. (Note that $(h', \tilde{a})$ is indeed a subhistory of h since i is not-yet-matched at h.) □

Lemma 3.5 *Fix a history h of M. If agent i is not-yet-matched at h, then $G(h) \subseteq O_i(h) \cup D_i^<(h)$. If i is a lurker at h then $G(h) \subseteq D_i^<(h)$.*

Proof of Lemma 3.5 It is enough to show that all houses in $O(h) \setminus (O_i(h) \cup D_i^<(h))$ have lurkers at h. Assume for contradiction that some $o \in O(h) \setminus (O_i(h) \cup D_i^<(h))$ has no lurker at h. By Lemma 3.4, for every not-yet-matched agent $j \neq i$, there exists $o_j \neq o$ with $R_j : o_j, \ldots, o \in \mathcal{R}_j(h)$. Since $o \notin D_i^<(h)$, Lemma 3.2 implies that $h \in Path(M(\mathbf{T}(R)))$ for $R_i : o$ and arbitrary $R_k \in \mathcal{R}_k(h)$ for all agents k already matched. Since $o \notin O_i(h)$ and since all other agents either are matched (not to o, since $o \in O(h)$) or prefer according to R all other houses over o, we therefore have that o is either unmatched or matched to an agent preferring all houses over o — a contradiction to Pareto optimality (since R_i ranks o highest).

Now, if i lurks some house o at h, then $O_i(h) = \{o\} \cup D_i^<(h)$ and $o \notin G(h)$. Combining this with the preceding paragraph, we obtain $G(h) \subseteq D_i^<(h)$ as desired. □

Lemma 3.6 *Let h be a history of M, let $o_l \in D_i(h)$ have a lurker and $o \in G(h)$. Let i be an active agent at h. Then either $o \in D_i(h)$ or i can force o at h.*

Proof of Lemma 3.6 Since $o_l \in D_i(h)$, some $l \neq i$ lurks o at h. It suffices to show that i can force any $o \in G(h) \setminus D_i(h)$. Since $o \in G(h) \setminus D_i(h)$, by Lemma 3.5 we have that $o \in O_i(h)$. Since $o \notin D_i(h)$ and $o_l \in D_i(h)$, we have that $o \neq o_l$. Lemma 3.5 and $o \in G(h)$ imply $o \in D_l(h)$. Since $o \in O_i(h) \setminus D_i(h)$, we have that $R_i : o, o_l \in \mathcal{R}_i(h)$. We will show that $M(\mathbf{T}(R))(i) = o$ holds for every R_{-i} s.t. $h \in Path(M(\mathbf{T}(R))$. Since l is a lurker for o_l at h, we have that $o_l = \max_{R_l} O_l(h)$. By Lemma 3.2, also $R'_l : o_l, o \in \mathcal{R}_i(h)$. Since $o_l \in D_i(h)$ and $o \in D_l(h)$, i and l respectively must R_i-prefer and R'_l-prefer their matches under $M(\mathbf{T}(R'_l, R_{-l}))$ to o_l and o. Since M is Pareto optimal, $M(\mathbf{T}(R'_l, R_{-l}))(i) = o$ and $M(\mathbf{T}(R'_l, R_{-l}))(l) = o_l$. Since M is strategy proof, $M(\mathbf{T}(R))(l) = o$ must also hold. Since $M(\mathbf{T}(R))(i) R_i o_l$ and $M(\mathbf{T}(R))(i) \neq o_l$ we have that $M(\mathbf{T}(R))(i) = o$, as required. □

Lemma 3.7 *Let h be a history of M, and let i be an active nonlurker at h. If for every $o \in G(h)$, either $o \in D_i(h)$ or i can force o for himself at h, then i is a dictator at h.*

Proof of Lemma 3.7 Assume that i is not a dictator at h. Therefore, there exists $o_l \in O_i(h) \setminus D_i(h)$ that i cannot force at h. By assumption, $o_l \notin G(h)$, and so o_l has a lurker, say l, at h. Since i is not a lurker, $l \neq i$.

If $P(h) \neq i$ let $h' = h$. Otherwise, let $h' = (h, \tilde{a})$ (which is well defined, given that $\tilde{a} \in \overline{A_h^*}$, since i is not a dictator). By gradual revelation $P(h') \neq i$, say $P(h') = k$. By Lemma 3.2, $Y_{h'}^* \neq \emptyset$, so there exists $o^* \in Y_{h'}^*$. We reason by cases.

Case 1: If $o^* \in G(h)$, then by assumption, either $o^* \in D_i(h)$ or i can force o^* at h. If $o^* \in D_i(h)$, Lemma 3.3(2) implies that $o^* \in O_i(h)$. Since l is a lurker at h and $o^* \in G(h)$, by Lemma 3.6, we have that $o^* \in D_l(h)$. Consider $R_l : o_l, o^*$ and either $R_i : o_l, o^*$ if $o^* \in D_i(h)$

or R_i that forces o^* from h if i can force o^* at h. Note that $R_l \in \mathcal{R}_l(h')$ and $R_i \in \mathcal{R}_i(h')$ in either case. By strategyproofness, i and l together get o_l and o^* under $M(\mathbf{T}(R))$ for every $R_{-i,l}$ s.t. $h' \in Path(M(\mathbf{T}(R)))$, and so k cannot force o^* at h' — a contradiction.

Case 2: If $o^* = o_l$, then let $o' \in G(h)$ and replace o^* in the above argument from Case 1 with o' to show that there exist $R_l \in \mathcal{R}_l(h')$ and $R_i \in \mathcal{R}_i(h')$ s.t. for every $R_{-i,l}$ with $h' \in Path(M(\mathbf{T}(R)))$, it is the case that i and l are matched with o' and o_l, implying once again that k cannot force $o_l = o^*$ at h' — a contradiction.

Case 3: Finally, if $o^* \notin G(h)$ and $o^* \neq o_l$, then o^* has a lurker, say l^*, at h. Let $o' \in G(h)$. Note that o^*, o_l, o' are all distinct. By the argument of Case 2, there exist $R_l \in \mathcal{R}_l(h')$ and $R_i \in \mathcal{R}_i(h')$ s.t. for every $R_{-i,l}$ with $h' \in Path(M(\mathbf{T}(R)))$, it is the case that i and l are matched with o' and o_l. Since $l*$ is a lurker at h and $o' \in G(h)$, by Lemma 3.6, we have that $o' \in D_l(h)$. Therefore, $R_{l^*} : o^*, o' \in \mathcal{R}_i(h')$ and by strategyproofness, for every $R_{-i,l,k}$ with $h' \in Path(M(\mathbf{T}(R)))$, l^* is matched with either o^* or o', but since i and l are matched with o' and o_l, in fact l^* is matched with o^*, implying once again that k cannot force o^* at h' — a contradiction. □

The next Lemma combines the two preceding ones to show that any agent i who has a lurked house o_l in his set $D_i(h)$ is a dictator at h.

Lemma 3.8 *Let h be a history of M. If $o_l \in D_i(h)$ has a lurker and if i is not a lurker, then i is a dictator at h.*

Proof of Lemma 3.8 Assume that i is not a dictator at h. By Lemma 3.6, for every $o \in G(h)$ either $o \in D_i(h)$ or i can force o at h. Since $o_l \in D_i(h)$ and since i is not a dictator at h, agent i is an active non-lurker at h, and by Lemma 3.7 i is a dictator — a contradiction. □

Lemma 3.9 *Let h be a history of M. Let $i_1, \ldots, i_L$ be all lurkers at h, ordered by the subhistory of h at which they turn into lurkers (i.e., i_1 became a lurker at a subhistory of the history at which i_2 became a lurker, etc.). For every $j \in \{1, \ldots, L\}$, let o_j be the house lurked by j.*

1. *$o_1, \ldots, o_L$ are distinct.*
2. *For every $j \in \{1, \ldots, L\}$, it is the case that $O_{i_j}(h) = O(h) \setminus \{o_1, \ldots, o_{j-1}\}$.*

Proof We prove the claim by induction. Say that $i_1, \ldots, i_{j-1}$ are the lurkers at a history h, ordered by the subhistory at which they became lurkers, say that they respectively lurk $o_1, \ldots, o_{j-1}$, and say that i_j becomes a lurker for a house o_j at h. (So $i_j = P(h)$, agent i_j is not a lurker at h, but is a lurker at $(h, \tilde{a})$.) Since i_j becomes a lurker for o_j at h, he is not a dictator at h.

Since $G(h) \subseteq O_i(h)$, we have that if o_j already has a lurker at h, then $G(h) \subseteq D_i(h)$, and so by Lemma 3.7, i_j is a dictator at h — a contradiction. Therefore, o_j is distinct from

$o_1, \ldots, o_{j-1}$. Since i not a dictator, then by Lemma 3.8, $D_i(h)$ contains no lurked agents. Since $O_i(h) = \{o_j\} \cup D_i(h)$, we therefore have that $O_i(h)$ contains no lurked agents. Note also that since i is not a dictator, we have By Lemmas 3.3(2) and 3.5 that $G(h) \subseteq O_i(h)$. Therefore, $O_i(h) = G(h) = O(h) \setminus \{o_1, \ldots, o_{j-1}\}$, as required. (And by monotonicity of D_i and O_i, the set O_i does not change in subsequent histories as long as i is a lurker.) $\square$

Lemma 3.10 *Let h be a history of M with lurked houses. Let h' be a maximal superhistory of h of the form $(h, \tilde{a}, \tilde{a}, \ldots, \tilde{a})$, and let $t = P(h')$. If $o \in O_k(h)$ for some nonlurker k and some house o that is lurked at h, then $k = t$. Moreover $D_t(h') = O(h)$.*

Proof of Lemma 3.10 Assume that there exists a nonlurker k with $o_l \in O_k(h) \setminus G(h)$. Let l be the lurker of o_l at h; by assumption, $l \neq k$. Assume for contradiction that $k \neq t$. Since $o_l \in O_k(h)$, then $M(\mathbf{T}(R))(k) = o_l$ for some R with $h \in \mathit{Path}(M(\mathbf{T}(R)))$.

If $h' \notin \mathit{Path}(M(\mathbf{T}(R)))$ then some agent i chooses an action $a \neq \tilde{a}$ before the maximal history h' is reached. Since any such agent is not a dictator, she can by Lemma 3.8 only choose a house o' that is not lurked at h. Since o' is not lurked at h, we have that $o' \in D_l(h)$ and l must be matched with a house she weakly R_l-prefers to o'. This holds in particular if $R_l : o_l, o'$ (since l lurks o_l, we have that $R_l \in \mathcal{R}_l(h)$), meaning that l must be matched with o_l if she declares R_l and i chooses a. By strategyproofness, and since l lurks o_l, we have that l must also be matched with o_l for any $R_l \in \mathcal{R}_l(h)$ if i chooses a. So the non-lurker k cannot be matched with o_l in this case. So we must have $h' \in \mathit{Path}(M(\mathbf{T}(R)))$.

Since h' is not terminal, let a be the unique action s.t. $(h', a) \in \mathit{Path}(M(\mathbf{T}(R)))$. Since $a \neq \tilde{a}$, agent t becomes matched to some house o at (h', a). If $o = o_l$, we are done. If $o \in D_l(h)$ then $l \neq k$ must, by the arguments given above, be matched with o_l. If not, then o must have an older (i.e., preceding in the order of Lemma 3.9) lurker than l. By strategyproofness, this older lurker, say l', is matched with some $o' \in D_{l'}(h)$. If $o' = o_l$, then we are done. If $o' \in D_l(h)$ then we are done as $l \neq k$ must by the arguments given above be matched with o_l. To avoid both scenarios, l' must be matched with a house with a lurker older than l, but younger than l'. Since there are only finitely many lurkers, we cannot go on indefinitely and o_l must be matched to one of the lurkers (either l or an older one) — a contradiction.

We now show that $D_t(h') = O(h)$. Let o_l be the house lurked by l, the oldest lurker at h (i.e., $l = i_1$ and $o_l = o_1$ in the notation of Lemma 3.9). Since l is a lurker, there is a profile of preference R with $h \in \mathit{Path}(M(\mathbf{T}(R)))$ and $M(\mathbf{T}(R))(l) \neq o_l$. Since by Lemma 3.9, we have that $o_l \notin O_{l'}(h)$ for every lurker l' other than l and since as shown above t is the only nonlurker with $o_l \in O_t(h)$, by the above argument it must be that $A(h')$ contains an action a with which t can force o_l, so $o_l \in D_t(h')$. Assume now that there exists a house $o' \in O(h) \setminus D_t(h')$. Since l is the oldest lurker in h and since $o' \neq o_l$, we have that $o' \in D_l(h)$. Since t is a dictator at h', we have that $o' \notin O_t(h')$. We note that $R_t : o', o_l \in \mathcal{R}_t(h')$ and $R_l : o_l, o' \in \mathcal{R}_l(h')$. For every $R_{-t,l}$ such that $h' \in \mathit{Path}(M(\mathbf{T}(R)))$, by strategyproofness, t is matched with o_l while l is matched with o' — a contradiction to Pareto optimality. $\square$

Lemma 3.11 *Let h be a history of M s.t. two distinct players i, j are active nonlurkers at h, and $k = P(h) \notin \{i, j\}$ has $D_k^<(h) = \emptyset$. Then, either i or j is a dictator at h.*

Proof of Lemma 3.11 Assume that all agents at h are either active or have not yet moved at h. (If not, then remove from M all players that have moved before h but are no longer active at h, along with their houses, keeping all else equal, to obtain a new incentive compatible gradual revelation mechanism for the smaller set of agents and houses.)

Assume for contradiction that neither i nor j is a dictator at h. Therefore, by Lemmas 3.3(2), 3.7 and 3.8, $D_i(h) \subsetneq G(h) \subseteq O_i(h)$ and $D_j(h) \subsetneq G(h) \subseteq O_j(h)$. Furthermore, by Lemma 3.10 at most one of the two agents can be matched with any lurked houses, so we have either $G(h) = O_i(h)$ have or $G(h) = O_j(h)$ (or both). By Lemma 3.2, and since the choice of k at h is nontrivial, there exists $o \in Y_h^* = D_k(h)$. We reason by cases.

Case 1: $D_i(h) \cup D_j(h) = G(h)$. Assume w.l.o.g. that $O_i(h) = G(h)$. We will obtain a contradiction by showing that j can force any $o_1 \in G(h) \setminus D_j(h)$ at h (and therefore, by Lemma 3.7, is a dictator). Since $o_1 \in G(h) \setminus D_j(h)$, we have that $o_1 \in D_i(h)$. Let $R_i \in \mathcal{R}_i(h)$. By Lemma 3.2, R_i ranks o_2 highest among $O_i(h) = G(h)$ for some $o_2 \in G(h) \setminus D_i(h)$. Therefore, $o_2 \in D_j(h)$. We first show that against $R_j : o_1, o_2$, agent i can force o_2 by playing $R'_i : o_2, o_1$. Indeed, by strategyproofness, in this case each agent gets either o_1 or o_2, and by Pareto optimality, agent i gets o_2. By strategyproofness for agent i, we therefore have that $\mathbf{T}_i(R_i)$ guarantees o_2 for agent i against $\mathbf{T}_j(R_j)$. We now claim that against $\mathbf{T}_i(R_i)$, agent j can force o_1 for himself by stating R_j (for R_j as defined above). Similarly to above, we have that each agent gets either o_1 or o_2, and as we have shown in this case that i gets o_2, we have that j gets o_1. As we have shown that for every $\mathbf{T}_i(R_i)$ that reaches h, agent j can force o_1 for himself against $\mathbf{T}_i(R_i)$ (using possibly different strategies for different such R_i), by strategyproofness we therefore have that agent j has a *single* strategy that forces o_1 against every such $\mathbf{T}_i(R_i)$; in other words, this strategy forces o_1 for agent j from h — a contradiction.

Assume henceforth, therefore, that $D_i(h) \cup D_j(h) \subsetneq G(h)$. By Lemma 3.5, we have that $O_k(h) \supseteq G(h)$. Assume for now that $o \in G(h)$.

Case 2: There exist $o_1 \in D_i(h) \setminus D_j(h)$, $o_2 \in D_j(h) \setminus D_i(h)$, and $o \in Y_h^* \subseteq G(h) \setminus (D_i(h) \cup D_j(h))$. Since M is obviously strategy proof, by an argument of Li (2015), it is obviously strategy proof on any restricted domain. We restrict the domain of preferences where agents i, j, k always rank o_1, o_2, o above all other houses and each other agent ℓ has a fixed preference $R_\ell \in \mathcal{R}_\ell(h)$ that has o_1, o_2, o at the bottom (this is possible since no agent lurks o_1, o_2, or o). Obtain a new mechanism by pruning all choices that are never taken. Since all agents other than i, j, and k have the same — known — preference in the restricted domain, none of these agents has any choice to make. After condensing all nodes that proved no choice we obtain a new (not necessarily gradual revelation) obviously strategy proof and Pareto optimal mechanism $\hat{M}$ in which only i, j, and k move. The history $\hat{h}$ in the new tree that corresponds to h in the original tree is such that $D_i(\hat{h}) = \{o_1\}$, $D_j(\hat{h}) = \{o_2\}$, and

$D_k(\hat{h}) = \{o\}$ and $O_i(\hat{h}) = O_j(\hat{h}) = O_k(\hat{h}) = \{o_1, o_2, o\}$.

Fix R with $R_i : o_2, o_1$, $R_j : o_1, o_2$ and $\hat{h} \in Path(\hat{M}(\mathbf{T}(R)))$. Since $D_i(\hat{h}) = \{o_1\}$, $D_j(\hat{h}) = \{o_2\}$, and since $\hat{M}$ is strategyproof and Pareto optimal, $\hat{M}(\mathbf{T}(R))(i) = o_1$ and $\hat{M}(\mathbf{T}(R))(j) = o_2$. This implies in particular that k cannot force o_1 at $\hat{h}$. Fix R' with $R'_i : o, o_2, o_1$ and $R'_j : o, o_2, o_1$ and $R'_k : o_1$. We have $M(\mathbf{T}(R'))(k) = o_1$ since $\hat{M}$ is Pareto optimal. Since k cannot force $o_1 \in O_k(\hat{h})$, but may prefer o_1 most, we have that k is not a dictator at $\hat{h}$ and has by Lemma 3.2 a unique action $\tilde{a} \in A(\hat{h})$ that does not determine his match.

Assume that $i = P(\hat{h}, \tilde{a})$ (the analysis for the cases $P(\hat{h}, \tilde{a}) \in \{j, k\}$ is analogous). By Lemma 3.2, $\mathcal{R}_i(\hat{h}, \tilde{a}) = \{R_i \mid \max_{R_i}\{o_1, o_2, o\} \neq o\}$. Again by Lemma 3.2, we have that agent i must have an action a that forces $o' \in \{o_2, o\}$ with $\mathcal{R}_i(\hat{h}, \tilde{a}, a) = \{R_i \mid \max_{R_i}\{o_1, o_2, o\} = o'\}$. To see that i cannot have a strategy that forces either o_2 or o, let $R_k : o_2, o, o_1$ and $R_j : o, o_2, o_1$ implying that k and j by strategyproofness are together matched to o_2 and o — no matter which strategy i uses.

Case 3: There exists $o^* \in D_i(h) \cap D_j(h)$, and $o \in G(h) \setminus (D_i(h) \cup D_j(h))$. Consider a profile R with $R_i : o, o^*$ and $R_j : o, o^*$. We note that any such profile reaches h. By strategyproofness, we have that i and j together receive o and o^* — a contradiction to the fact that k can force o at h.

Assume henceforth, therefore, that $o \in D_i(h) \cup D_j(h)$.

Case 4: $o \in D_i(h) \cap D_j(h)$. Recall that there exists $o' \in G(h) \setminus (D_i(h) \cup D_j(h))$. Consider a profile R with $R_i : o', o$ and $R_j : o', o$. We note that any such profile reaches h. By strategyproofness, we have that i and j together receive o' and o — a contradiction to the fact that k can force o at h.

Case 5: $o \in D_i(h) \setminus D_j(h)$ and there exists $o_2 \in D_j(h) \setminus D_i(h)$. (The case $o \in D_j(h) \setminus D_i(h)$ and there exists $o_1 \in D_i(h) \setminus D_j(h)$ is analyzed analogously.)

Consider a profile R with $R_i : o_2, o$ and $R_j : o, o_2$. We note any such profile reaches h. By strategyproofness, we have that i and j together receive o_2 and o — a contradiction to the fact that k can force o at h.

Case 6: $o \in D_i(h) \setminus D_j(h)$ and, $D_i(h) \supseteq D_j(h) \cup Y^*_h$, and there exists $o_2 \in D_j(h) \setminus Y^*_h$. (The case $o \in D_j(h) \setminus D_i(h)$, $D_j(h) \supseteq D_i(h) \cup Y^*_h$ and there exists $o_1 \in D_i(h) \setminus Y^*_h$ is analyzed analogously.) We will show that i can force any $o' \in G(h) \setminus D_i(h)$ via some strategy from h, therefore, by Lemma 3.7, obtaining a contradiction to the assumption that i is not a dictator at h.

Assume therefore, for contradiction, that there exists a strategy profile R_{-i} that reaches h s.t. for no $R_i \in \mathcal{R}_i(h)$ does i get o'. Let i play the following strategy: always pass (i.e., $\tilde{a}$) unless he can ensure o' via an action a with $Y(a) = \{o'\}$. By assumption, at some history along the path dictated by R_{-i} and this strategy, it would no longer hold that $o' \in O_i$. We claim that as long as j and k have not specifically played an action other than pass (i.e., $\tilde{a}$), this cannot be. Indeed, as long as none of i, j, k indicate that it is not the case that they all prefer o, o', o_2 the most, then it is possible that they do prefer these houses the most. Under such preferences, agents i and k get "at least" o and agent j gets "at least" o_2; therefore, they will all together get these three items. If at any history h' along the way, one of i, j plays an action that ensures for her some item $o'' \in D_i(h')$, then since i's strategy up until that point is consistent with playing as if he likes o' most and o'' second, then by continuing to play according to these preferences, agent i ensures o' by strategyproofness. It therefore suffices to show that at no history h' along this path can j or k play an action that ensures an item $o'' \notin D_i(h')$. Indeed, otherwise let h' be such a minimal history and w.l.o.g. $j = P(h')$. By Lemmas 3.6 and 3.7, we can assume w.l.o.g. that $o'' \in G(h')$. We obtain a contradiction as in Case 5: since the play so far of i and k is consistent with the preferences that rank o'' most (since $o'' \notin D_i(h')$ and by minimality of h' also $o'' \notin D_k(h')$ since $D_k(h') \subseteq D_i(h')$) and rank $o \in D_i(h) \cap D_k(h) \subseteq D_i(h') \cap D_k(h')$ second, then by strategyproofness they will together get o'' and o, and so j cannot ensure o'' for himself at h'.

Finally, we consider the case in which $o \notin G(h)$. Recall that $D_k^<(h) = \emptyset$. By definition, there is a lurker $\ell \neq k$ for o at h, and so by Lemma 3.6, k can force any house $o_1 \in G(h) = O_i(h)$ from h, and so by the analysis of either Case 4 or Case 5 we are done. □

Lemma 3.12 *Let h be a history of M with lurked houses and let t be as in Lemma 3.10. If $k = P(h) \neq t$ is a nonlurker (by Lemma 3.11, no other active nonlurkers exist except t and k) and t is not a dictator at h, then $Y_h^* \cap D_t(h) = \emptyset$.*

Proof of Lemma 3.12 Let $o' \in D_t(h)$. We will show that $o' \notin Y_h^*$. Let l be the oldest lurker at h, and let o_l be her lurked house (i.e., $l = i_1$ and $o_l = o_1$ in the notation of Lemma 3.9). Since t is not a dictator at h, we have by Lemma 3.8 that $o_l \notin D_t(h)$, and so $o_l \neq o'$. Since l is the oldest lurker at h, we have that $o' \in O(h) \setminus \{o_l\} = D_l(h)$. By Lemmas 3.10 and 3.3(2), $o_l \in O_t(h)$. Let $R_l : o_l, o'$ and $R_t : o_l, o'$. Since $o' \in D_t(h) \cap D_l(h)$ while $o_l \in (O_l(h) \setminus D_l(h)) \cap (O_t(h) \setminus D_t(h))$, we have by strategyproofness that when playing R_l and R_t (which reach h) from h, l and t get o_l and o' for themselves, and so k cannot force o' at h and so $o' \notin Y_h^*$, as required. □

The characterization is a direct consequence of the above lemmas: since if an agent i is a dictator at some history h, then we can have him "divulge" his entire preferences at that point, and thus he becomes inactive, and therefore there always exists a mechanism implementing the same social choice function where no "already-active" dictators are still active while others make nondictatorial moves (or become active).

As long as two active agents i, j are nonlurkers, then no one else can join the game and i and j slowly accumulate houses into D_i and D_j, respectively. If one of them, say i, becomes a lurker, then her lurked house becomes impossible for all agents except one, which we denote by t (for terminator), and from that point no agent can accumulate houses that are in D_t. Since i became a lurker, another player k can join the game, and if one of k and j becomes a lurker, then another can join the game, etc. At some point, the terminator t will be a dictator, and after her choice, all lurkers by strategyproofness will become matched via a dictatorship from the "oldest" lurker i (who has the largest list of possible houses) to the "youngest" lurker (who has the smallest list of possible houses), and no lurkers remain. If no one forced a house from the option list D_ℓ of the nonlurker that did not claim a house, then he continues (along with D_ℓ) to the next round.

3.1 Matching with Outside Options

When agents may prefer being unmatched over being matched to certain houses (and possibly also more agents exist than houses), the OSP-implementable and Pareto optimal social choice functions are similar in spirit, but considerably more detailed to describe, due to the following phenomenon: Recall that at any point when an agent becomes a lurker, the sets of possible houses O_i of some other agents i are reduced. In this case, such agents i may be asked by the mechanism whether they prefer being unmatched over being matched with any house in (the reduced) O_i (note that for this reason, the set O_i has to be tracked and updated for all agents and not only for active agent $i \in T$). If such an agent indeed prefers being unmatched, then she can divulge her full preferences, and based on these preferences, the mechanism may now reduce O_i for some other agent, etc. (thus, in a sense, dynamically choosing the "terminator" agent t defined in the proof above), or even award some lurkers their lurked houses. The analysis for this case is similar, yet more cluttered, and we omit it.

4 Combinatorial Auctions with Additive Bidders

In a combinatorial auction with additive bidders, there are $m \geq 0$ goods and $n > 1$ agents, called bidders. Each bidder has a nonnegative integer valuation for each of the goods, and the valuation of a bidder for a subset of the goods is simply the sum of that bidder's valuations for the goods in the subset. In such a setting, an outcome is the allocation of each good to some bidder and a specification of how much to charge each bidder, and the utility of each bidder from an outcome is his valuation of the subset of the goods that is awarded to him, minus the payment he is charged (bidders prefer higher utility over lower utility[16]). We will be interested in social choice functions with the following properties:

- Welfare maximizing: each good is awarded to a bidder who values it most.
- Losers pay nothing: bidders who are awarded no good pay 0.
- OSP-implementable.

[16]We note that the valuations of a bidder for all items are fully determined by the preferences of that bidder over outcomes. Indeed, a bidder values a good a by value v iff he is indifferent between getting nothing and being charged nothing, and getting a and being charged v.

Li (2015) shows that if $m = 1$, then an ascending-price implementation of a second-price auction satisfies the above properties. We will now show that this is as far as these properties can be stretched in combinatorial auctions, i.e., that for $m > 1$, no social choice function satisfies the above properties. In particular, even VCG with the Clarke pivot rule (Vickrey, 1961; Clarke, 1971; Groves, 1973), even when implemented as two ascending auctions, either simultaneously or consecutively, does not satisfy the above properties. (Due to discreteness of the valuation space, it is well known that there are other incentive compatible social choice functions beside VCG with the Clarke pivot rule that maximize welfare and in which losers pay nothing.) Furthermore, any setting with richer type sets (as long as they include additive valuations) has no social choice function that satisfies the above properties. It is enough to show this for $m = 2$ and $n = 2$, as this is a special case of any case with more goods and/or more bidders.

Theorem 3 *For $m = 2$ and $n = 2$, no social choice function satisfies the above properties.*

Proof Assume the contrary. Call the goods a and b and the bidders 1 and 2. We denote by $(v(a), v(b))$ the preferences induced by a valuation of $v(a)$ for good a and $v(b)$ for good b.

We first consider the possible outcomes when the preferences of each bidder belong to the set $\{(20, 2), (8, 8), (6, 6)\}$. (For ease of presentation, we make no attempt to choose particularly low valuations or to tighten the analysis of utilities below.)

If both bidders have the same preferences $(v(a), v(b))$, then we claim that the utility of each bidder is at most 2. (Under VCG with the Clarke pivot rule, the utility of each bidder would be 0.) Since in this case any allocation is welfare-maximizing, we reason by cases.

- If bidder 1 is awarded both goods and bidder 2 is awarded no good, then bidder 2 is charged nothing. Assume for contradiction that bidder 1 is charged less than $v(a) + v(b) - 2$. Then, bidder i with preferences $(v(a) - 1, v(b) - 1)$ (with the preferences of bidder 2 unchanged), who would get no good if declaring his true preferences (and thus have utility 0), would rather misrepresent himself as $(v(a), v(b))$ and get both goods (and have strictly positive utility), in contradiction to strategyproofness. So, the utility of bidder 1 is at most $v(a) + v(b) - (v(a) + v(b) - 2) = 2$. The case in which bidder 2 is awarded both goods and bidder 1 is awarded no good is analyzed analogously.

- If bidder 1 is awarded a and bidder 2 is awarded b, then assume for contradiction that bidder 1 is charged less than $v(a)-1$. In this case, bidder 1 with preferences $(v(a)-1, 0)$ (with the preferences of bidder 2 unchanged), who would get no good if declaring his true preferences (and thus have utility 0), would rather misrepresent himself as $(v(a), v(b))$ and get good a (and have strictly positive utility), in contradiction to strategyproofness. So, the utility of bidder 1 is at most $v(a)-(v(a)-1) = 1$. Similarly, bidder 2 is charged at least $v(b) - 1$ and so has utility at most 1. The case in which bidder 2 is awarded a and bidder 1 is awarded b is is analyzed analogously.

If bidder 1 has preferences $(20, 2)$ and bidder 2 has preferences $(8, 8)$, then bidder 1 gets good a and bidder 2 gets good b. In this case, bidder 1 is charged at most 9. Indeed, to see this note that if bidder 1 had preferences $(9, 2)$ (with the preferences of bidder 2 unchanged), then he would also get good a, but by strategyproofness would be charged at most 9 as otherwise his utility would be negative and he would have rather misrepresented his preferences to be $(0, 0)$ for a utility of 0. Therefore, since bidder 1 with preferences $(20, 2)$ has no incentive to misrepresent his preferences as $(9, 2)$, he is charged at most 9 as well. Similarly, bidder 2 is charged at most 3. (Under VCG with the Clarke pivot rule, they would be charged 8 and 2, respectively.) Similarly, if bidder 1 has preferences $(8, 8)$ and bidder 2 has preferences $(20, 2)$, then bidder 1 is charged at most 3 and bidder 2 is charged at most 9.

If bidder 1 has preferences $(20, 2)$ and bidder 2 has preferences $(6, 6)$, then bidder 1 gets good a and bidder 2 gets good b. In this case, similarly to the above analysis, bidder 1 is charged at most 7 and bidder 2 is charged at most 3. (Under VCG with the Clarke pivot rule, they would be charged 6 and 2, respectively.) Similarly, if bidder 1 has preferences $(6, 6)$ and bidder 2 has preferences $(20, 2)$, then bidder 1 is charged at most 3 and bidder 2 is charged at most 7.

Finally, If bidder 1 has preferences $(8, 8)$ and bidder 2 has preferences $(6, 6)$, then bidder 1 gets both goods and bidder 2 gets no good (and is charged nothing). In this case, bidder 1 is charged at most 14. Indeed, to see this note that if bidder 1 had preferences $(7, 7)$ (with the preferences of bidder 2 unchanged), then he would also get both goods, but by strategyproofness would be charged at most 14 as otherwise his utility would be negative and he would have rather misrepresented his preferences to be $(0, 0)$ for a utility of 0. Therefore, since bidder 1 with preferences $(8, 8)$ has no incentive to misrepresent his preferences as $(7, 7)$, he is charged at most 14 as well. (Under VCG with the Clarke pivot rule, bidder 1 would be charged 12.) Similarly, if bidder 1 has preferences $(6, 6)$ and bidder 2 has preferences $(8, 8)$, then bidder 1 is charged nothing and bidder 2 is charged at most 14.

To see that the given social choice function cannot be implemented by a strategyproof mechanism, suppose it could, and let $\hat{M}$ be an obviously incentive compatible gradual revelation mechanism that implements it. We now prune $\hat{M}$ by restricting the preferences of both bidders to only be in the set $\{(20, 2), (8, 8), (6, 6)\}$, to obtain an obviously incentive compatible gradual revelation mechanism M for the restricted preference domain.

Assume without loss of generality that $P(\emptyset) = 1$. Since the choice of bidder 1 at $\emptyset$ is real (i.e., more than one action exists), then since bidder 1 has only three possible preferences, some action a is chosen by a single preference of bidder 1, and so reveals his preferences (any other preference for bidder 1 chooses an action different from a). We reach a contradiction by reason by cases.

If action a is chosen by the preferences $(20, 2)$, then the minimal utility bidder 1 can get by reporting truthfully is at most 2 (when bidder 2 has preferences $(20, 2)$ as well). Nonetheless, the maximal utility bidder 1 can get by deviating by misreporting his preferences as $(8, 8)$ is at least $22 - 14 = 8 > 2$ (when bidder 2 has preferences $(6, 6)$). This contradicts obvious strategyproofness.

If action a is chosen by the preferences $(8,8)$, then the minimal utility bidder 1 can get by reporting truthfully is at most 2 (when bidder 2 has preferences $(8,8)$ as well). Nonetheless, the maximal utility bidder 1 can get by deviating by misreporting his preferences as $(6,6)$ is at least $8-3=5>2$ (when bidder 2 has preferences $(20,2)$). This contradicts obvious strategyproofness.

Finally, if action a is chosen by the preferences $(6,6)$, then the minimal utility bidder 1 can get by reporting truthfully is at most 2 (when bidder 2 has preferences $(6,6)$ as well). Nonetheless, the maximal utility bidder 1 can get by deviating by misreporting his preferences as $(8,8)$ is at least $6-3=3>2$ (when bidder 2 has preferences $(20,2)$). This contradicts obvious strategyproofness. □

5 Voting

Majority voting is not obviously strategyproof even when there are just two options. In fact, exactly one supermajority rule is obviously strategyproof: The rule for which y^* is chosen if and only if all agents prefer y^* over y' is obviously strategyproof. In the sequential implementation of any other supermajority rule, the first agent $P(\emptyset)$ does not have an action that determines one of the two choices. So for whichever choice she picks, the worst-case scenario is that all other agents vote against her. On the other hand, the best-case scenario for the option she did not choose is that all other agents will vote for her preferred option.

There are, however, some nondictatorial obviously strategyproof unanimous voting mechanisms: Any proto-dictatorship with only two options is obviously strategyproof. Note that M is a proto-dictatorship for $Y:=\{y,z\}$ if at each nonterminal history h (precisely) one of the following hold:

- $Y_h^*=\{y\}$ and $\overline{A_h^*}=\{\tilde{a}\}$ with $Y(\tilde{a})=\{y,z\}$.
- $Y_h^*=\{z\}$ and $\overline{A_h^*}=\{a\}$ with $Y(\tilde{a})=\{y,z\}$, or
- $Y_h^*=\{y,z\}$ (and $\overline{A_h^*}=\emptyset$),

and if $P(h')\neq P(h)$ for any two distinct nonterminal histories h,h' s.t. h is a subhistory of h'.

Theorem 4 *Let $Y=\{y,z\}$. Then M is obviously strategyproof and onto if and only if it is a proto-dictatorship.*

Proof Fix any social choice function scf that is implementable via an obviously strategyproof mechanism. By Theorem 1, scf must be implementable by an obviously incentive compatible gradual revelation mechanism M. Let h,i be such that h is a minimal history with $P(h)=i$. Since M is a gradual revelation mechanism, i must have at least two choices at h (i.e., $|A(h)| \geq 2$). Since there are only two possible preferences for i and since M is a gradual revelation mechanism, there are at most $2 = |\mathcal{R}_i|$ choices for i at h. In sum, we have $|A(h)| = 2$. Moreover, there exists no h' with $h \subsetneq h'$ and $P(h') = i$, since i already fully reveals his preference at h. By Lemma 2.1, $Y^*_h \neq \emptyset$. So, h must be covered by one of the three above cases.

To see that any proto-dictatorship is obviously strategyproof, it is enough to analyze histories h in which $Y^*_h = \{y\}$ and $\overline{A^*_h} = \{\tilde{a}\}$ with $Y(\tilde{a}) = \{y, z\}$ (histories h with $Y^*_h = \{z\}$ are analyzed analogously, and in histories h with $Y^*_h = \{y, z\}$, the choosing agent is a dictator). In this case, $P(h)$ ensures that the outcome is y if y is his preferred option. If z is his preferred option, then choosing $\tilde{a}$ is obviously strategyproof: the best outcome under the deviation to ensuring y is identical to the worst outcome given $\tilde{a}$. □

6 Single Peaked Preferences

In the domain of single peaked preferences, the possible outcomes (also called policies) are $Y = \mathbb{Z}$, and each agent (also called voter) has single peaked preferences, i.e., prefers some $y \in \mathbb{Z}$ the most, and for every $y'' > y' \geq y$ or $y'' < y' \leq y$, strictly prefers y' over y''. A unanimous social choice rule is one that, if the ideal points of all agents coincide, chooses the joint ideal point. (Unanimity is a strictly weaker assumption than Pareto optimality.)

With single peaked preferences, it is well known that there is a range of strategyproof and unanimous social choice functions. Most prominently, median voting, which maps any profile of preferences to a median of the set of all ideal points is strategyproof and unanimous (and even Pareto optimal). However, median voting is not obviously strategy proof when there are at least 3 voters. To see this, suppose some gradual direct revelation mechanism did implement median voting. Say the most preferred choices of the first agent, i.e., $P(\emptyset)$, in this mechanism is y, and let $y' \in Y \setminus \{y\}$. Say that truthtelling prescribes for this agent to choose some action $a \in A(\emptyset)$. If all other voters say their ideal point is y', then regardless of the preferences of the first agent, the median of all announced preferences, i.e., y', is implemented. If the the first agent deviates to some action $a' \neq a$ and if all voters — according to the best possible case scenario — say their ideal point is y, then the median of all announced preferences, i.e., y, is implemented. In sum, truthtelling is not obviously strategyproof for voter 1.

A different, less popular, unanimous (and even Pareto optimal) and strategyproof social choice function for any single peaked domain is the function *min*, which maps any profile of preferences to the minimal ideal point. If the set of possible ideal points is bounded from below by some bound $\underline{y}$, then this function is also OSP-implementable: The obviously strategyproof implementation of *min* follows along the lines of the (obviously strategyproof) implementation of second price auctions via ascending bid auctions. The *min* mechanism starts with $\underline{y}$. For each number $y \in [\underline{y}, \infty) \cap \mathbb{Z}$, sequentially in increasing order, each agent is given an option to decide whether to continue or to stop. When one agent stops at some y, the mechanism terminates with the social choice y.

Assume that the ideal point of some agent i is y^*. We claim that continuing at any $y < y^*$ and then stopping at y^* is obviously strategyproof. Indeed, continuing at any $y < y^*$ is obviously strategyproof, since the worst-case outcome under the strategy that continues until y^* and then stops at y^* is in $[y, y^*] \cap \mathbb{Z}$, and therefore no worse than y, while the best possible outcome when deviating to stopping at y is y. Stopping at $y = y^*$ is obviously strategyproof as it implements i's top choice. By the same argument, the social choice function *max*, which maps any profile of preferences to the maximal ideal point is OSP-implementable if the set of possible ideal points is bounded from above by some bound $\overline{y}$.

As we show in this section, the set of obviously strategyproof and unanimous mechanisms are, in a precise sense, combinations of *max*, *min* and dictatorship, and furthermore, they are all Pareto optimal. **Dictatorship with safeguards against extremism** are defined for domains of single peaked preferences. There is one agent, say 1, who is called the dictator. All other agents have limited veto rights. Specifically, each agent $i \neq 1$ can block extreme leftists policies and rightist policies in the rays $(-\infty, l^i)$ and (r^i, ∞), for some $l_i \leq r_i \in \mathbb{Z} \cup \{-\infty, \infty\}$. Furthermore, there exists some y^m with $l^i \leq y^m \leq r^i$ for all i. Say that $\underline{y}^i$ and $\overline{y}^i$ respectively are agent i's preferred policies in the rays $(-\infty, l^i]$ and $[r^i, \infty)$. Then the outcome of the dictatorship with safeguard against extremism is agent 1's most preferred policy in $\bigcap_{i \neq 1} [\overline{y}^i, \underline{y}^i]$.

According to this social choice function, agent 1 is free to choose any policy "in the middle": If 1's ideal policy y is in $[\max_{i \neq 1} l^i, \min_{i \neq 1} r^i]$, then y is implemented. Note that by assumption, $y^m \in [\max_{i \neq 1} l^i, \min_{i \neq 1} r^i]$, and so this choice set is nonempty. If agent 1's ideal policy is further to the left or right, then it may only be chosen if none of a select group of citizens vetoes this choice. As we consider more extreme policies, the group that needs to consent to the implementation of a policy increases. Dictatorships with safeguards against extremism embed standard dictatorships ($l^i = -\infty$ and $r^i = \infty$ for all i). They also embed *min* when the ideal points are bounded from below by some $\underline{y}$ (by $r^i = \underline{y}$ for all i) as well as *max* when the ideal points are bounded from above by some $\overline{y}$ (by $l^i = \overline{y}$ for all i).

Fix a dictatorship with safeguards against extremism scf. Then scf is implementable in obviously dominant strategies by the mechanism that first offers the dictator to choose any option in $\bigcap_{i \neq 1} [\overline{y}^i, \underline{y}^i] \colon = [L^*, H^*]$. If the dictator does not choose an option in this interval, he indicates whether the mechanism is to continue to the left or to the right (according to the direction of the dictator's ideal point). If the mechanism continues to the right, then similarly to the implementation of *min*, the mechanism starts with H^* and for each number $y \in [H^*, \infty)$, sequentially in increasing order, each agent i with either $i = 1$ (the dictator) or $r^i \leq y$ is given an option to decide whether to continue or to stop at y. When one agent stops at some y, the mechanism terminates with the social choice y. Similarly, if the mechanism continues to the left, then similarly to the implementation of *max*, the mechanism starts with L^* and for each policy $y \in (-\infty, L^*]$, the dictator (agent 1) and each agent i with $l^i \geq y$ may decide to stop the mechanism with the implementation of y.

To see that the mechanism is Pareto optimal, first consider the case where the dictator chooses a policy y from $[L^*, H^*]$. In this case, the dictator strictly prefers y to all other policies and the outcome is Pareto optimal. If the dictator initiates a move to (say) the right, then the mechanism either stops at the ideal policy of some agent, or it stops at a policy that is to the left of the dictator's ideal point and to the right of the ideal point of the agent who chose to stop. In either case, Pareto optimality is satisfied.

Remark 3 *If we only demand for any agent with ideal point $y \in \mathbb{Z}$, that for every $y'' > y' \geq y$ or $y'' < y' \leq y$, this agent weakly prefers y' over y'', then after someone says "stop" at some value y, to ensure Pareto optimality, the mechanism would start going in the other direction until some agent (who was allowed to say stop w.r.t. y) says stop again, which such a player does not do as long as he is indifferent between the current value and the one that will follow it.*

If we require our social choice function to cover finer and finer grids in $\mathbb{R}$, then only the above mechanism is OSP-implementable. However, with our fixed grid, namely $Y = \mathbb{Z}$, the set of obviously strategyproof and unanimous mechanisms is slightly larger than the set of dictatorships with safeguards against extremism. We may then combine dictatorships with safeguards against extremism with the proto-dictatorships of Theorem 4. When such a mechanism moves to the right or left, some agents may in addition to stopping or continuing at y call for "arbitration" between y and a directly neighboring option. More specifically, if, e.g., the mechanism goes right to y', some specific agent with $r_i = y' + 1$ may not only force that the outcome is at most $y' + 1$, but also to choose to initiate an "arbitration" between y' and $y' + 1$ via a proto-dictatorship (whose parameters depend on $y' + 1$).[17] In such a case, the obviously strategyproof implementation allows i the choice between forcing $y' + 1$, initiating an arbitration, and continuing, immediately after all relevant players were given the option to stop at y' and before any other player is given the option to stop at $y' + 1$.

Theorem 5 *If $Y = \mathbb{Z}$, a social choice function scf for the domain of single peaked preferences is unanimous and obviously strategyproof implementable if and only if it is a dictatorship with safeguards against extremism. Moreover, in this case scf is also Pareto optimal.*

We prove Theorem 5 via Lemmas 6.1 through 6.8.

Lemma 6.1 *Any dictatorship with safeguards against extremism is Pareto optimal (and in particular unanimous) and obviously strategyproof implementable.*

Proof As outlined above. □

For Lemmas 6.2 through 6.8, fix an obviously incentive compatible gradual revelation mechanism M that implements a given social choice function, with the following property:

[17]Similarly, if the set of ideal points is bounded from (say) above by some $\overline{y}$, then one specific agent with $r_i = \overline{y}$ may choose arbitration between $\overline{y} - 1$ and $\overline{y}$.

For each nonterminal history h of M, there does not exist another obviously incentive compatible gradual revelation mechanism M' that implements the same social choice function as M and such that M and M' coincide except for the subtree at h, and such that $|\overline{A^*_h}|$ is at most 1 at M, but greater than 1 at M'. Such an M always exists: start with any obviously incentive compatible gradual revelation mechanism M' that implements the given social choice function, and then, considering first each nodes h in the first level in the tree of M', if h violates the above condition, replace the subtree at h with another subtree that satisfies the above condition. Next replace all subtrees that violate the above condition at nodes in the second level, then in the third, and so forth. Since each node does not change any more after some finite number of steps (equal to the level of this node), the resulting mechanism is well defined, even though the height of the tree of M (and so the number of steps in the process defining M) may be infinite.

Fix an obviously incentive compatible gradual revelation mechanism M that implements a unanimous social choice function. Assume that M satisfies the above property and assume w.l.o.g. that $P(\emptyset) = 1$.

Lemma 6.2 *Let h be a nonterminal history in M and let $i = P(h)$. Let $y \in Y$ s.t. $y \in Y(h, a)$ and $y+1 \in Y(h, a')$ for two distinct $a, a' \in A(h)$. If there exist*[18] *$R_i : y, y+1 \in \mathcal{R}_i(h)$ and $R'_i : y+1 \in \mathcal{R}_i(h)$ (or $R_i : y \in \mathcal{R}_i(h)$ and $R'_i : y+1, y \in \mathcal{R}_i(h)$), then $\{y, y+1\} \cap Y^*_h \neq \emptyset$.*

Proof We prove the lemma for the first case (swapping y and $y+1$ obtains the proof for the second case). Suppose that $y, y+1 \notin Y^*_h$. Assume w.l.o.g. that $\mathbf{T}_i(R_i)(h) = a$. By Lemma 2.1 and by definition of $R_i, R'_i \in \mathcal{R}_i(h)$, we have that $y, y+1 \notin Y(h, a) \cap Y(h, a')$, so $y \notin Y(h, a')$ and $y+1 \notin Y(h, a)$. Since $y \notin Y^*_h$, there is some preference profile R_{-i} for all other agents with $h \in Path(\mathbf{T}(R))$ and $M(\mathbf{T}(R)) = y' \neq y$. Since the second ranked choice under R_i, namely $y+1$, is not in $Y(h, a)$, we have $y' \neq y+1$. Since $y+1 \in Y(h, a')$, there exists a preference profile R' s.t. $h \in Path(\mathbf{T}(R'))$, $\mathbf{T}_i(R'_i)(h) = a'$ and $M(\mathbf{T}(R)) = y+1$. A contradiction to the obvious strategyproofness arises, since $y+1 = M(\mathbf{T}(R'))P_i M(\mathbf{T}(R)) = y'$. □

,

Lemma 6.3 *Let $i \in N$ and let h be a minimal nonterminal history in M s.t. $P(h) = i$. If $Y(h) = Y$, then $Y^*_h \neq \emptyset$.*

Proof Suppose $Y^*_h = \emptyset$. Since M is a gradual revelation mechanism, $A(\emptyset)$ must contain at least two choices. So, there exists $y \in Y$ such that $y \in Y(h, a)$, $y+1 \in Y(h, a')$ for $a \neq a'$ and $y, y+1 \notin Y^*_h$ (and recall that $\mathcal{R}_i(h) = \mathcal{R}_i$), a contradiction to Lemma 6.2. So Y^*_h must be nonempty. □

Lemma 6.4 *The set $Y^*_\emptyset$ is a nonempty. There exist numbers $L^* \in \{-\infty\} \cup \mathbb{Z}$ and $H^* \in \mathbb{Z} \cup \{\infty\}$ s.t. $L^* \leq H^*$ and $Y^*_\emptyset = [L^*, H^*] \cap \mathbb{Z}$*

[18] Recall that we use, e.g., $R_i : y, y+1$ to denote a preference R_i for agent i that ranks y first and $y+1$ second.

Proof By Lemma 6.3, $Y^*_\emptyset \neq \emptyset$. Let $y^* < y^\circ$ be two policies in $Y^*_\emptyset$. Suppose we had $y' \in (y^*, y^\circ) \cap Y$ but $y' \notin Y^*_\emptyset$. Since $y' \notin Y^*_\emptyset$, there exists a preference profile R such that R_1 ranks y' at the top but the outcome of the mechanism is $M(\mathbf{T}(R)) = \tilde{y} \neq y'$. Assume without loss of generality that $\tilde{y} < y'$.

Define two preference profiles R' and R''_{-1} such that $R'_i : \tilde{y}$ and $R''_i : y'$ for all $i \neq 1$, and such that $R'_1 : y'$ and R'_1 ranks y° strictly above $\tilde{y}$. Starting with the profile R and inductively switching the preference of each agent $i \neq 1$ from R_i to R'_i, the strategyproofness of M implies that $M(\mathbf{T}(R_1, R'_{-1})) = \tilde{y}$. Since R'_1 ranks y° strictly above $\tilde{y}$, since $y^\circ \in Y^*_\emptyset$, since R'_1 is single peaked, and since M is strategyproof, we have that $M(\mathbf{T}(R')) \in (\tilde{y}, y^\circ]$.

Assume for contradiction that $M(\mathbf{T}(R')) > y'$. Since $\tilde{y} < y'$, since $R'_2 : \tilde{y}$, and since by strategyproofness for we have (similarly to he above argument for R') that $M(\mathbf{T}(R''_2, R'_{-2})) > \tilde{y}$, we have that $M(\mathbf{T}(R''_2, R'_{-2})) \geq M(\mathbf{T}(R'))$ holds by strategyproofness. (Indeed, if we had $M(\mathbf{T}(R')) > M(\mathbf{T}(R''_2, R'_{-2})) > \tilde{y}$ then agent 2 with preference R'_2 would have an incentive to lie.) Inductively switching the preference of each agent $i \neq 1$ from R'_i to R''_i and applying the preceding argument, we obtain that $M(\mathbf{T}(R'_1, R''_{-1})) \geq M(\mathbf{T}(R'))$. Since $M(\mathbf{T}(R')) > y'$ we obtain a contradiction to unanimity, which requires $M(\mathbf{T}(R'_1, R''_{-1})) = y'$ as all preferences in (R'_1, R'_{-1}) have the ideal point y'. Therefore, $M(\mathbf{T}(R')) \leq y'$.

Therefore, $M(\mathbf{T}(R_1, R'_{-1})) = \tilde{y} < M(\mathbf{T}(R')) \leq y'$, and so $M(\mathbf{T}(R'))P_1M(\mathbf{T}(R_1, R'_{-1}))$, contradicting the strategyproofness of M. So we must have $y' \in Y^*_\emptyset$, and therefore $y' \in Y^*_\emptyset$ is a nonempty "interval". □

For Lemmas 6.5 through 6.7, let $Y^*_\emptyset : = [L^*, H^*]$ with $L^* \leq H^*$. If $L^* = -\infty$ and $H^* = \infty$ then M is a dictatorship and we are done. So assume that either $L^* > -\infty$ or $H^* < \infty$ (or both).

Lemma 6.5 *Let h be a nonterminal history. If $\overline{A^*_h} = \{a\}$ for some action a, then for every $y^* \in Y^*_h$, either $y^* \leq Y(a)$ or $y^* \geq Y(a)$.*

Proof Let $i = P(h)$. Assume for contradiction that $X = \{y \in Y(h,a) \mid y < y^*\}$ and $Z = \{y \in Y(h,a) \mid y > y^*\}$ are both nonempty for some $y^* \in Y^*_h$. Let

$$Y_X = \{R_i \in \mathcal{R}_i(h) \mid \mathbf{T}_i(R_i)(h) = a \ \ \& \ \text{ the peak of } R_i \text{ is} \leq y^*\}$$

$$Y_Z = \{R_i \in \mathcal{R}_i(h) \mid \mathbf{T}_i(R_i)(h) = a \ \ \& \ \text{ the peak of } R_i \text{ is} > y^*\}$$

We claim that for every $R_i \in Y_X$, for every R_{-i} s.t. $h \in Path(\mathbf{T}(R))$, it is the case that $M(\mathbf{T}(R)) \in X \cup \{y^*\}$. Indeed, if $M(\mathbf{T}(R)) \in Z$, then we would have a contradiction to strategyproofness because at h agent i can ensure that the outcome is y^*, which he prefers over $M(\mathbf{T}(R)) > y^*$. Similarly, for every $R_i \in Y_Z$, for every R_{-i} s.t. $h \in Path(\mathbf{T}(R))$, it is the case that $M(\mathbf{T}(R)) \in Z \cup \{y^*\}$. Since X and Z are nonempty, Y_X and Y_Z are nonempty as well. (Otherwise we would have a contradiction to the fact that $Y(a)$ contains the disjoint union of X and Y, and possibly also y^*.)

Define two new two new distinct actions l and r. Let $\mathcal{R}_i(l) = Y_X$ and let l lead to a copy of the subtree that a leads to, where all choices outside $\mathcal{R}_i(l)$ have been pruned. Similarly Let $\mathcal{R}_i(r) = Y_Z$ and let r lead to a copy of the subtree that a leads to, where all choices outside $\mathcal{R}_i(r)$ have been pruned. Define a new tree by replacing $a \in A(h)$ with the two new actions l and r. If this new mechanism $\hat{M}$ is obviously incentive compatible (under the above-defined truthtelling strategy), then it implements the same social choice function as the original mechanism, and so our assumption that if $|\overline{A^*_h}| = 1$ then no modification can result in $|\overline{A^*_h}| > 1$ is violated. (Note that indeed we have in $\hat{M}$ that $l, r \in \overline{A^*_h}$ and so $|\overline{A^*_h}| = 2$, since by gradual revelation any preference R_i that chooses a in the original mechanism does not ensure the final outcome, and so also every preference R_i that chooses l or r in $\hat{M}$ does not ensure the final outcome.)

To see that the new mechanism $\hat{M}$ is indeed obviously incentive compatible (under the above-defined truthtelling strategy), note that for every history h' that is not a subhistory of h, the conditions for obvious strategyproofness do not change, while for every history h' that is a strict subhistory of h, when a player makes a choice, he faces a pruned version of the subtree he faced in the original mechanism, and therefore obvious strategyproofness is maintained. It remains to show that obvious strategyproofness is not violated for i at h. Indeed, obvious strategyproofness when playing any action in A^*_h is maintained since i would — under the original mechanism and under $\hat{M}$ — for any deviation at best get his best choice from $Y(h)$, which has not changed. Similarly, obvious strategyproofness when playing l or r is maintained w.r.t. deviating to forcing some outcome, because the worst outcome when playing l or r in $\hat{M}$ is no worse than the worst outcome when playing a in the original mechanism, and deviation was not incentivized there. Finally, obvious strategyproofness when playing l is maintained w.r.t. deviating to playing r, since the best outcome that a type that chooses l (and by definition has peak $\leq y^*$) can get by playing r is no better than y^* (since all possible outcomes when playing r are in $Z \cup \{y^*\}$ and therefore are $\geq y^*$), however the worst outcome such a type gets when playing l is $\leq y^*$. Similarly, obvious strategyproofness when playing r is maintained w.r.t. deviating to playing l as well. □

Lemma 6.6 *Following are all actions in* $\overline{A^*_\emptyset}$.

1. *If* $H^* < \infty$, *then* $\overline{A^*_\emptyset}$ *contains an action* r *with* $\mathcal{R}_i(r) = \{R_i :$ *ideal point of* $R_i > H^*\}$ *and* $Y(r) = [H^*, \infty) \cap \mathbb{Z}$.
2. *If* $-\infty < L^*$, *then* $\overline{A^*_\emptyset}$ *contains an action* l *with* $\mathcal{R}_i(l) = \{R_i :$ *ideal point of* $R_i < L^*\}$ *and* $Y(l) = (-\infty, L^*] \cap \mathbb{Z}$.

Proof Assume that $H^* < \infty$ (the proof when $H^* = \infty$ and $L^* > -\infty$ is analogous to the proof when $H^* < \infty$ and $L^* = -\infty$). Assume for contradiction that there exist two different actions $a, a' \in A(\emptyset)$ such that $Y(a) \cap [H^* + 1, \infty) \neq \emptyset \neq Y(a') \cap [H^* + 1, \infty)$. So there must exist some $y > H^*$ such that $y \in Y(a)$ and $y + 1 \in Y(a')$, a contradiction to Lemma 6.2. Therefore there is at most one action $r \in A(\emptyset)$ such that $Y(r) \cap [H^* + 1, \infty) \neq \emptyset$. By unanimity, $\bigcup_{a\in A(\emptyset)} Y(a) = Y$, and so for each $y \in [H^*+1,\infty)$ there must exist an action a such that $y \in Y(a)$. By the preceding two statements, there exists a unique action $r \in A(\emptyset)$ s.t. $Y(r) \cap [H^*+1,\infty) \neq \emptyset$, and furthermore, $Y(r) \supseteq [H^*+1,\infty) \cap \mathbb{Z}$.

Assume for contradiction that $H^* \notin Y(r)$. Consider the preference profile R_1 that ranks $H^*{+}1$ first, and ranks H^* second. By unanimity, $\mathbf{T}_1(R_1)(\emptyset) = r$ (as no other action has $H^*{+}1$ as a possible outcome), however, since $H^*+1 \notin Y^*_\emptyset$, we have that there exists a preference profile R_{-i} s.t. $h \in Path(\mathbf{T}(R))$ and s.t. $M(\mathbf{T}(R)) \neq H^*+1$. Therefore, since $H^* \in Y^*_h$, we obtain a contradiction to obvious strategyproofness. Therefore, $Y(r) \supseteq [H^*,\infty) \cap \mathbb{Z}$.

Mutatis mutandis if $L^* > -\infty$. there exists a unique action $l \in A(\emptyset)$ s.t. $Y(l) \cap (-\infty, L^*-1] \neq \emptyset$, and furthermore, $Y(l) \supseteq (-\infty, L^*] \cap \mathbb{Z}$. We note that we have not yet determined whether or not $l = r$.

To complete the proof, we reason by cases. Assume first that either $L^* = -\infty$ or $r \neq l$. Therefore, if $L^* > -\infty$, then $Y(r) \cap (-\infty, L^*-1] = \emptyset$, and so $Y(r) \subseteq [L^*,\infty) \cap \mathbb{Z}$. As regardless of whether or not $L^* > -\infty$, we have in the current case that $Y(r) \subseteq [L^*,\infty) \cap \mathbb{Z}$, to complete the proof for this case, it is enough to show that $Y(r) \cap [L^*, H^*-1] = \emptyset$. Assume for contradiction that there exists $y \in Y(r) \cap [L^*, H^*-1]$. Therefore, there exists a preference profile R s.t. $\mathbf{T}_1(R_1)(\emptyset) = r$ and $M(\mathbf{T}(R)) = y$. Since the mechanism is a gradual revelation mechanism, there exists $y' \in Y(r) \setminus \{y\}$ and a preference profile R'_{-1} s.t. $M(\mathbf{T}(R_1, R'_{-1})) = y'$. By strategyproofness and since $y \in Y^*_\emptyset$, we have that $y' R_1 y$. If $y R_1 y'$ as well, then the peak y'' of R_1 is between y and y'; by strategyproofness, therefore $y'' \in Y(r) \setminus Y^*_\emptyset$ and so $y'' > H^*$. If $y' P_1 y$, then by strategyproofness, $y' \notin Y^*_\emptyset$; therefore, as $y' \in Y(r)$, we have that $y' > H^*$. Either way, there exists $y'' > H^*$ that R_1 ranks strictly above y. Since $H^* > y$, we therefore obtain a contradiction to strategyproofness, as $H^* \in Y^*_\emptyset$ is preferred by R_1 over $M(\mathbf{T}(R)) = y$.

It remains to consider the case in which $L^* > -\infty$ (recall also that $H^* < \infty$) and $r = l$, but such a setting is impossible, as it contradicts Lemma 6.5. □

Lemma 6.7 *Let h' and $h = (h, a')$ be two consecutive histories of M, and let $F \in \mathbb{Z}$ s.t. $\max Y^*_{h'} = F$ and s.t. for every $i \in N$, it is the case that $\mathcal{R}_i(h)$ contains (not necessarily exclusively) all preferences with peak $> F$. If $Y(h) = [F,\infty) \cap \mathbb{Z}$, then (precisely) one of the following holds:*

1. *$Y^*_h = \{F\}$ and $\overline{A^*_h} = \{a\}$ with $Y(a) = [F,\infty) \cap \mathbb{Z}$,*
2. *$Y^*_h = \{F, F+1\}$ and $\overline{A^*_h} = \{a\}$ with $Y(a) = [F+1,\infty) \cap \mathbb{Z}$, or*
3. *$Y^*_h = \{F+1\}$ and $\overline{A^*_h} = \{a, b\}$ with $Y(b) = \{F, F+1\}$ and $Y(a) = [F+1,\infty) \cap \mathbb{Z}$.*

Proof We first claim that no action a at h has $Y(a) \subseteq [F+2,\infty)$ (in particular, $Y^*_h \cap [F+2,\infty) = \emptyset$). Indeed, since the preferences of $P(h')$ might be s.t. the peak is $F+1$ and the second-best option is F (which $P(h')$ can ensure at h'), we would have a contradiction to strategyproofness if $P(h)$ could ensures that the outcome is no less than $F+2$. By

Lemma 6.2 and since $Y(h)$ contains all integer from some point, we have that since there is more than one action at h, then $Y^*_h \neq \emptyset$. Therefore, $\emptyset \neq Y^*_h \subseteq \{F, F+1\}$. We reason by cases.

If $Y^*_h = \{F\}$, then by Lemma 6.2, we have that precisely one action a exists with $Y(a) \cap [F+1, \infty) \neq \emptyset$. Therefore, $\overline{A^*_h} = \{a\}$, and furthermore $Y(a) = [F, \infty) \cap \mathbb{Z}$ or $Y(a) = [F+1, \infty) \cap \mathbb{Z}$, and it remains for this case to show that $F \in Y(a)$. Assume for contradiction that $F \notin Y(a)$. Consider the preference profile R_i that ranks $F+1$ first, and ranks F second. By unanimity, $\mathbf{T}_i(R_i)(\emptyset) = a$ (as no other action has $F+1$ as a possible outcome), however, since $F+1 \notin Y^*_h$, we have that there exists a preference profile R_{-i} s.t. $h \in Path(\mathbf{T}(R))$ and s.t. $M(\mathbf{T}(R)) \neq F+1$. Therefore, since $F \in Y^*_h$, we obtain a contradiction to obvious strategyproofness. Therefore, $Y(a) = [F, \infty) \cap \mathbb{Z}$, as required.

We are left with the cases in which Y^*_h is either $\{F, F+1\}$ or $\{F+1\}$. By Lemma 6.2, we have that an action a exists with $Y(a) \supseteq [F+2, \infty) \cap \mathbb{Z}$, and that $Y(b) \cap [F+2, \infty) = \emptyset$ for any other action b. Since we have shown that no action can ensure an outcome of at least $F+2$. We claim that $F+1 \in Y(a)$. Indeed, since the preferences of $P(h)$ might be s.t. the peak is $F+2$ and second is $F+1$, then by unanimity $P(h)$ should choose $Y(a)$ under such preferences, however since he cannot ensure $F+2$, and since he can force $F+1$ be deviation, if $F+1 \notin Y(a)$ we would have a contradiction to strategyproofness. So, $Y(a) \supseteq [F+1, \infty) \cap \mathbb{Z}$.

Assume first that $Y^*_h = \{F, F+1\}$. Assume for contradiction that $F \in Y(a)$. Therefore, there exists $R_i \in \mathcal{R}_i(h)$ s.t. there exists R_{-i} s.t. $h \in Path(\mathbf{T}(R))$ and $M(\mathbf{R}(T)) = F$. Since $P(h)$ can force $F+1$ at h, by strategyproofness we therefore have that R_i has ideal point $\leq F$. Therefore, since $F \in Y^*_h$ and by strategyproofness, R_i forces the outcome from h to be F, and so by gradual revelation does not choose $a \in \overline{A^*_h}$ — a contradiction. Therefore, $F \notin Y(a)$, and so $Y(a) = [F+1, \infty) \cap \mathbb{Z}$. Since as any action $b \neq a$ at h has $Y(b) \cap [F+2, \infty) = \emptyset$, it has $Y(b) \subseteq \{F, F+1\} = Y^*_h$, and so by gradual revelation and strategyproofness, $b \in A^*_h$. Therefore, $\overline{A^*_h} = \{a\}$, as required.

Finally, assume that $Y^*_h = \{F+1\}$. Since $F \in Y^*_h$, there exists $b \in \overline{A^*_h}$ s.t. $F \in Y(b)$. Furthermore, and since $P(h)$ can ensure the outcome to be $F+1$ at h, by strategyproofness every $R_i \in \mathcal{R}_i(h)$ that chooses b at h has peak $\leq F$. Therefore, by strategyproofness and since $F \notin Y^*_h$, we have by Lemma 2.1 that b is the unique action at h s.t. $F \in Y(b)$. Therefore, any action $c \neq \{a, b\}$ at h has $Y(c) = F+1$, and so $A^*_h = \{a, b\}$. If $a = b$, then by Lemma 6.5 we obtain a contradiction, and so $a \neq b$. Therefore, $Y(a) = [F+1, \infty) \cap \mathbb{Z}$ and $Y(b) = \{F, F+1\}$, as required. □

Similarly, we obtain the following lemma (a "mirror version" of Lemma 6.7):

Lemma 6.8 *Let h' and $h = (h, a')$ be two consecutive histories of M, and let $F \in \mathbb{Z}$ s.t. $\min Y^*_{h'} = F$ and s.t. for every $i \in N$, it is the case that $\mathcal{R}_i(h)$ contains (not necessarily exclusively) all preferences with peak $< F$. If $Y(h) = (-\infty, F] \cap \mathbb{Z}$, then (precisely) one of the following holds:*

1. *$Y^*_h = \{F\}$ and $\overline{A^*_h} = \{a\}$ with $Y(a) = (-\infty, F] \cap \mathbb{Z}$,*

2. $Y^*_h = \{F, F-1\}$ *and* $\overline{A^*_h} = \{a\}$ *with* $Y(a) = (-\infty, F-1] \cap \mathbb{Z}$, *or*

3. $Y^*_h = \{F-1\}$ *and* $\overline{A^*_h} = \{a, b\}$ *with* $Y(b) = \{F, F-1\}$ *and* $Y(a) = (-\infty, F-1] \cap \mathbb{Z}$,

The characterization follows from Lemmas 6.6 through 6.8: By Lemma 6.6, if the dictator is not happy with any option he can force, then he chooses l (left) or right (r), according to where his ideal point lies. Assume w.l.o.g. that he chooses to go right. Then initialize $F = H^*$, and by Lemma 6.7 (if he chooses to go left, then Lemma 6.8 is used), some other player chooses is given one of the following options.

1. Option 1: Force F, Option 2: continue, where F (and everything higher) is still "on the table".

2. Option 1: Force F, Option 2: force $F+1$, Option 3: continue, where only $F+1$ (and everything higher) is on the table.

3. Option 1: Force F, Option 2: restrict to $F, F+1$ ("arbitrate", from here must start an OSP mechanism that chooses between these two options, i.e., a proto-dictatorship), Option 3: continue, where only $F+1$ (and everything higher) is on the table.

If this agent chooses to keep F on the table, then some other player is given the same choice w.r.t. F, etc. Otherwise, F is incremented by one and some other player is given the same choice w.r.t. the "new" F, etc.

For any $F \geq H^*$, let D_F be the set of players that were given the option to force F as outcome. (D_{H^*} includes the dictator by definition.) We claim that D_F is nondecreasing in F. Indeed, for any player who was given the option to force the outcome to be F but not to force the outcome to be $F+1$, we have a contradiction w.r.t. the preferences that prefer $F+1$ the most and F second, as by unanimity he cannot force F, but therefore he may end up with $F+2$ or higher. Finally, note that for any given F, only one player can choose to arbitrate between F and $F+1$, and since that player can force $F+1$ at that point, by strategyproofness it follows that he was not given the option to force F before that, and so the history at which he was allowed to choose "arbitrate" was the first history at which he was given any choice.

Acknowledgments

This collaboration is supported by the ARCHES Prize. Yannai Gonczarowski is supported by the Adams Fellowship Program of the Israel Academy of Sciences and Humanities; his work is supported by ISF grants 230/10 and 1435/14 administered by the Israeli Academy of Sciences, and by Israel-USA Bi-national Science Foundation (BSF) grant number 2014389. We thank Sergiu Hart, Noam Nisan, and Ahuva Mu'alem for useful discussions and comments.

References

I. Ashlagi and Y. A. Gonczarowski. Stable matching mechanisms are not obviously strategy-proof. Mimeo, 2016.

A. Bogomolnaia, R. Deb, and L. Ehlers. Strategy-proof assignment on the full preference domain. *Journal of Economic Theory*, 123(2):161–186, 2005.

E. H. Clarke. Multipart pricing of public goods. *Public Choice*, 11(1):17–33, 1971.

A. Gibbard. Manipulation of voting schemes: a general result. *Econometrica: journal of the Econometric Society*, 41(4):587–601, 1973.

T. Groves. Incentives in teams. *Econometrica: Journal of the Econometric Society*, 41(4): 617–631, 1973.

S. Li. Obviously strategy-proof mechanisms. Mimeo, 2015.

M. Pycia. Obvious dominance and random priority. Mimeo, 2016.

M. A. Satterthwaite. Strategy-proofness and arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10(2):187–217, 1975.

L. Shapley and H. Scarf. On cores and indivisibility. *Journal of Mathematical Economics*, 1 (1):23–37, 1974.

P. Troyan. Obviously strategyproof implementation of allocation mechanisms. Private communication, 2016.

W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of Finance*, 16(1):8–37, 1961.