Red Nacional de Investigadores en Economía

Trade-off between Manipulability and Dictatorial Power: a Proof of the Gibbard–Satterthwaite Theorem

Agustín G. Bonifacio (UNSL/CONICET)

DOCUMENTO DE TRABAJO N° 254

Junio de 2023

Citar como:
Bonifacio, Agustín G. (2023). Trade-off between manipulability and dictatorial power: a proof of the Gibbard–Satterthwaite Theorem. *Documento de trabajo RedNIE N°254.*

Trade-off between manipulability and dictatorial power: a proof of the Gibbard–Satterthwaite Theorem*

Agustín G. Bonifacio†

June 15, 2023

Abstract

By endowing the class of *tops-only* and *efficient* social choice rules with a dual order structure that exploits the trade-off between different degrees of manipulability and dictatorial power rules allow agents to have, we provide a proof of the Gibbard–Satterthwaite Theorem.

JEL classification: D71, D72.

Keywords: Gibbard–Satterthwaite Theorem, manipulability, dictatorial power, tops-only rules.

1 Introduction

The Gibbard–Satterthwaite Theorem states that, when more than two alternatives and all possible preferences over alternatives are considered, a social choice rule is unanimous and strategy-proof if and only if it is dictatorial (Gibbard, 1973; Satterthwaite, 1975). There are several interesting proofs of the Gibbard–Satterthwaite Theorem (see Section 3.3 in Barberà, 2011, and references therein). In this paper, we provide a new proof of this theorem. To do this, we exploit the trade-off between the different degrees of manipulability

*I thank Pablo Arribillaga, Jordi Massó, and Pablo Neme for their helpful comments. I acknowledge financial support from UNSL through grants 032016, 030120, and 030320, from Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) through grant PIP 112-200801-00655, and from Agencia Nacional de Promoción Científica y Tecnológica through grant PICT 2017-2355.

†Instituto de Matemática Aplicada San Luis (UNSL and CONICET) and Departamento de Matemática, Universidad Nacional de San Luis, San Luis, Argentina. E-mail: abonifacio@unsl.edu.ar

and the different degrees of dictatorial power that rules allow agents to have within the class of tops-only and efficient rules.

The idea behind our proof is the following. First, we show that each unanimous and strategy-proof rule is both tops-only and efficient. Since dictatorial rules are tops-only and efficient as well, it is safe to restrain our analysis to this class of rules in order to prove Gibbard–Satterthwaite's result.

The novelty of our approach consists of defining two orders within the class of tops-only and efficient rules. One order compares rules according to their manipulability. A rule is at least as manipulable as another rule if the former has as many manipulable (preference) profiles as the latter.[1] Therefore, since strategy-proof rules have no manipulable profiles, any rule is at least as manipulable as a strategy-proof rule. The other order compares rules according to the dictatorial power of the agents. A rule is at least as dictatorial as another rule if the former has as many dictatorial profiles as the latter, where a dictatorial profile is such that agents not obtaining their top alternative cannot unilaterally affect the social outcome by changing their preference. Therefore, since dictatorial rules have all of their profiles dictatorial, any dictatorial rule is at least as dictatorial as any other rule.

The crucial fact is that, given a tops-only and efficient rule, each preference profile is either manipulable or dictatorial for that rule, which is equivalent to saying that both orders are dual. Gibbard–Satterthwaite's result follows easily from this. Clearly, every dictatorial rule is strategy-proof. To see the converse, consider a strategy-proof rule. As we already pointed out, any other rule is at least as manipulable as this strategy-proof rule. Then, by the duality between the orders, this strategy-proof rule is at least as dictatorial as any other rule. Thus, this strategy-proof rule is dictatorial.

The paper is organized as follows. After the preliminaries are presented in Section 2, the class of tops-only and efficient rules is introduced in Section 3, where it is also shown that every unanimous and strategy-proof rule belongs to the class. Finally, in Section 4, the dual order structure of tops-only and efficient rules together with the proof of the Gibbard–Satterthwaite Theorem are presented.

2 Preliminaries

A set of *agents* $N = \{1, \ldots, n\}$, with $|N| \geq 2$, has to choose an alternative from a finite set X, with $|X| \geq 3$. Each agent $i \in N$ has a strict *preference* P_i over X. Denote by $t(P_i)$ to the best alternative according to P_i, called the *top* of P_i. Sometimes we write $P_i : x, y, z, \ldots$, meaning that x is the top of P_i, y is the second-best alternative of P_i, z the third-best,

[1] Maus et al. (2007) compare rules according to this criterion as well. Other similar criteria are studied, for example, in Arribillaga and Massó (2016) and Pathak and Sönmez (2013).

etc. Given $i \in N$ and $x \in X$, a generic preference for i with a top equal to x is denoted by P_i^x. We denote by R_i the weak preference over X associated to P_i. Let $\mathcal{P}$ be the set of all strict preferences over X. A *(preference) profile* is an ordered list of n preferences, $P = (P_1, \ldots, P_n) \in \mathcal{P}^n$, one for each agent. Given a profile P and a set of agents S, P_{-S} denotes the subprofile in $\mathcal{P}^{n-|S|}$ obtained by deleting each P_i for $i \in S$ from P.

A *(social choice) rule* is a function $f : \mathcal{P}^n \longrightarrow X$ that selects an alternative in X for each preference profile in $\mathcal{P}^n$. We assume throughout that rules are *unanimous*, i.e., for each $P \in \mathcal{P}^n$ such that $t(P_i) = x$ for each $i \in N$, $f(P) = x$. Given a rule $f : \mathcal{P}^n \longrightarrow X$, a profile $P \in \mathcal{P}^n$ and a preference $P_i' \in \mathcal{P}$, we say that *agent i manipulates f at P via P_i'* if $f(P_i', P_{-i}) P_i f(P)$. If no agent ever manipulates f, then f is *strategy-proof*. The set of all *strategy-proof* and *unanimous* rules is denoted by $\mathcal{S}$. A rule $f : \mathcal{P}^n \longrightarrow X$ is *dictatorial* if there is $i \in N$ (the *dictator*) such that, for each $P \in \mathcal{P}^n$, $f(P) = t(P_i)$. The set of all *dictatorial* rules is denoted by $\mathcal{D}$. The Gibbard–Satterthwaite Theorem states that $\mathcal{S} = \mathcal{D}$.

3 Tops-only and efficient rules

A rule $f : \mathcal{P}^n \longrightarrow X$ is *tops-only* if for each $P, P' \in \mathcal{P}^n$ such that $t(P_i) = t(P_i')$ for each $i \in N$, $f(P) = f(P')$. A (seemingly) weaker property is the following. A rule $f : \mathcal{P}^n \longrightarrow X$ is *own-top-only* if for each $P \in \mathcal{P}^n$ and each $P_i' \in \mathcal{P}$ such that $t(P_i') = t(P_i)$, $f(P_i', P_{-i}) = f(P)$. It is easy to see that both properties are equivalent.

A rule $f : \mathcal{P}^n \longrightarrow X$ is *efficient* if, for each $P \in \mathcal{P}^n$, there is no $x \in X$ such that $x P_i f(P)$ for each $i \in N$. Let $\mathcal{T}$ denote the set of all *tops-only* and *efficient* rules. It is clear that *dictatorial* rules are *efficient* and *tops-only*, i.e., $\mathcal{D} \subseteq \mathcal{T}$. Next, we prove that *unanimous* and *strategy-proof* rules are *efficient* and *tops-only* as well.

Lemma 1 *If $f \in \mathcal{S}$, then f is efficient.*

Proof. Let $f \in \mathcal{S}$. Assume that it is not *efficient*. Then, there are $x \in X$ and $P \in \mathcal{P}^n$ such that $x P_i f(P)$ for each $i \in N$. Let $P^\star \in \mathcal{P}^n$ be such that, for each $i \in N$, $P_i^\star : x, f(P), \ldots$ By *strategy-proofness*, $f(P_1^\star, P_{-1}) \neq x$ (otherwise agent 1 manipulates f at P via $P_1^\star$). Furthermore, again by *strategy-proofness*, $f(P_i^\star, P_{-i}) = f(P)$ (otherwise, since $f(P_i^\star, P_{-i}) \neq x$, agent 1 manipulates f at $(P_i^\star, P_{-i})$ via P_i). Using the same argument, changing the preference of one agent at a time, it follows that $f(P_{-n}^\star, P_n) = f(P)$. By *unanimity*, $f(P^\star) = x$. Therefore, $f(P^\star) P_n f(P_{-n}^\star, P_n)$, contradicting *strategy-proofness*. Thus, f is efficient. □

To see that any rule in $\mathcal{S}$ is *tops-only*, we first present two auxiliary results.

Lemma 2 *Let $f \in \mathcal{S}$, $P \in \mathcal{P}^n$, and $i \in N$ be such that $t(P_i) \neq f(P)$ and let $x \in X \setminus \{t(P_i), f(P)\}$. If $P_i' \in \mathcal{P}$ is such that $P_i' : t(P_i), x, f(P), \ldots$, then $f(P_i', P_{-i}) \in \{x, f(P)\}$.*

Proof. Let $f \in \mathcal{S}$, $P \in \mathcal{P}^n$, and $i \in N$ be as stated. Let $x \in X \setminus \{t(P_i), f(P)\}$ and $P_i' \in \mathcal{P}$ be such that $P_i' : t(P_i), x, f(P), \ldots$ Then, $f(P_i', P_{-i}) \in \{t(P_i), x, f(P)\}$ (otherwise agent i manipulates f at (P_i', P_{-i}) via P_i). Moreover, $f(P_i', P_{-i}) \neq t(P_i)$ (otherwise agent i manipulates f at P via P_i'). Therefore, $f(P_i', P_{-i}) \in \{x, f(P)\}$. $\square$

The following lemma says that rules in $\mathcal{S}$ always select one of the top alternatives in each profile of preferences.[2] Given $r \in X$ and $P \in \mathcal{P}^n$, let $N_r(P) \equiv \{i \in N : t(P_i)\} = r\}$.

Lemma 3 *Let $f \in \mathcal{S}$, $P \in \mathcal{P}^n$ and $x \in X$ be such that $f(P) = x$. Then, $N_x(P) \neq \emptyset$.*

Proof. Let $f \in \mathcal{S}$, $P \in \mathcal{P}^n$ and $x \in X$ be such that $f(P) = x$. Assume $N_x(P) = \emptyset$. Let $z = t(P_1)$ and let $k = \min\{i \in N \setminus \{1\} : t(\overline{P}_i) \neq z\}$. Such k exists by *unanimity* since $f(P) \neq z$. Let $w = t(P_k)$. Let $\overline{P}_1 \in \mathcal{P}$ be such that $\overline{P}_1 : z, w, x, \ldots$, and let $\overline{P}_k \in \mathcal{P}$ be such that $\overline{P}_k : w, z, x, \ldots$ By Lemma 2, $f(\overline{P}_k, P_{-k}) \in \{x, z\}$ and $f(\overline{P}_1, P_{-1}) \in \{x, w\}$. There are two cases to consider:

1. $\boldsymbol{f(\overline{P}_k, P_{-k}) = z}$. Then, $f(\overline{P}_{1,k}, P_{-1,k}) = z$ (otherwise agent 1 manipulates f at $(\overline{P}_{1,k}, P_{-1,k})$ via P_1). This, in turn, implies $f(\overline{P}_1, P_{-1}) \neq w$ (otherwise agent k manipulates f at $(\overline{P}_{1,k}, P_{-1,k})$ via P_k), so $f(\overline{P}_1, P_{-1}) = x$. Next, starting from $f(\overline{P}_1, P_{-1}) = x$, by *strategy-proofness* change the preferences of agents 2 to $k-1$ one at a time to obtain $f(\overline{P}_1, \ldots, \overline{P}_1, P_{-1,2,\ldots,k-1}) = x$. Then, by *strategy-proofness*, $f(\overline{P}_k, \overline{P}_1, \ldots, \overline{P}_1, P_{-1,2,\ldots,k-1}) \in \{x, w, z\}$ (otherwise agent 1 manipulates f at $(\overline{P}_k, \overline{P}_1, \ldots, \overline{P}_1, P_{-1,2,\ldots,k-1})$ via $\overline{P}_1$). Again by *strategy-proofness*, $f(\overline{P}_k, \overline{P}_1, \ldots, \overline{P}_1, P_{-1,2,\ldots,k-1}) \notin \{w, z\}$ (otherwise agent 1 manipulates f at $(\overline{P}_1, \overline{P}_1, \ldots, \overline{P}_1, P_{-1,2,\ldots,k-1})$ via $\overline{P}_k$), so $f(\overline{P}_k, \overline{P}_1, \ldots, \overline{P}_1, P_{-1,2,\ldots,k-1}) = x$. Changing the preferences of agents 2 to $k-1$ one at the time, using *strategy-proofness*, we get
$$f(\overline{P}_k, \overline{P}_k, \ldots, \overline{P}_k, P_{-1,2,\ldots,k-1}) = x. \tag{1}$$

2. $\boldsymbol{f(\overline{P}_k, P_{-k}) = x}$. Then, by *strategy-proofness*, $f(P_1, \ldots, P_{k-1}, \overline{P}_1, P_{k+1}, \ldots, P_n) \in \{x, w, z\}$ (otherwise agent k manipulates f at $(P_1, \ldots, P_{k-1}, \overline{P}_1, P_{k+1}, \ldots, P_n)$ via P_k). Also by *strategy-proofness*, $f(P_1, \ldots, P_{k-1}, \overline{P}_1, P_{k+1}, \ldots, P_n) \notin \{w, z\}$ (otherwise agent k manipulates f at $(\overline{P}_k, P_{-k})$ via $\overline{P}_1$). Thus,
$$f(P_1, \ldots, P_{k-1}, \overline{P}_1, P_{k+1}, \ldots, P_n) = x. \tag{2}$$

Notice that, starting from profile P in which the first $k-1$ tops are the same (equal to z) and different from $x = f(P)$, we construct in each case a new profile (see (1) and (2), respectively) in which the first k tops are the same and different from x. Repeating this argument $n-k$ times we violate *unanimity*. Therefore, $N_x(P) \neq \emptyset$. $\square$

[2]Our proof of this result is similar to the proof of Theorem 2 (a) in Ninjbat (2012).

Lemma 4 *If $f \in \mathcal{S}$, then f is tops-only.*

Proof. It suffices to show that f is *own-top-only*. Assume it is not. Then, w.l.o.g. there are $P_1', P_1^\star \in \mathcal{P}$ and $\overline{P}_{-1} \in \mathcal{P}^{n-1}$ such that $t(P_1') = t(P_1^\star)$ and $f(P_1', \overline{P}_{-1}) \neq f(P_1^\star, \overline{P}_{-1})$. Let

$$x \equiv f(P_1', \overline{P}_{-1}),\ y \equiv f(P_1^\star, \overline{P}_{-1}),\ \text{and } z \equiv t(P_1'). \tag{3}$$

Notice that all x, y, z are different. There are two cases to consider:

Case 1: $n = 2$. By Lemma 3, $t(\overline{P}_2) = x$. Consider $\widehat{P}_1 \in \mathcal{P}$ such that $\widehat{P}_1 : z, y, x, \ldots$ Then, by Lemma 2, $f(\widehat{P}_1, \overline{P}_2) \in \{x, y\}$. Moreover, $f(\widehat{P}_1, \overline{P}_2) \neq x$ (otherwise agent 1 manipulates f at $(\widehat{P}_1, \overline{P}_2)$ via $P_1^\star$). Therefore,

$$f(\widehat{P}_1, \overline{P}_2) = y. \tag{4}$$

Now, consider $\widetilde{P}_2 \in \mathcal{P}$ such that $\widetilde{P}_2 : x, z, y, \ldots$ Then, $f(P_1', \widetilde{P}_2) = x$ (otherwise agent 2 manipulates f at $(\widehat{P}_1, \widetilde{P}_2)$ via $\overline{P}_2$). Moreover, by Lemma 2, $f(\widehat{P}_1, \widetilde{P}_2) \in \{y, z\}$. Also, $f(\widehat{P}_1, \widetilde{P}_2) \neq z$ (otherwise agent 1 manipulates f at $(P_1', \widetilde{P}_2)$ via $\widehat{P}_1$). Therefore,

$$f(\widehat{P}_1, \widetilde{P}_2) = y. \tag{5}$$

To finish, consider any $P_2^z \in \mathcal{P}$. By *unanimity* and (5), $f(\widehat{P}_1, P_2^z) = z \widetilde{P}_2 y = f(\widehat{P}_1, \widetilde{P}_2)$, contradicting *strategy-proofness*. This implies that $f : \mathcal{P}^2 \longrightarrow X$ is *own-top-only*.

Case 2: $n \geq 3$. By Lemma 3, $N_x(P_1', \overline{P}_{-1}) \neq \emptyset$. Assume, w.l.o.g., that $N_x(P_1', \overline{P}_{-1}) = \{2, 3, \ldots, k\}$. Consider $P_2^x, P_3^x, \ldots, P_k^x \in \mathcal{P}$. By (3) and *strategy-proofness*,

$$x = f(P_1', \overline{P}_{-1}) = f(P_1', P_2^x, \overline{P}_{-1,2}) = \ldots = f(P_1', P_2^x, \ldots, P_k^x, \overline{P}_{-1,2,\ldots,k}). \tag{6}$$

Define a two-agent rule $g : \mathcal{P}^2 \longrightarrow X$ as follows. For each $(P_1, P_2) \in \mathcal{P}^2$,

$$g(P_1, P_2) \equiv f(P_1, \underbrace{P_2, \ldots, P_2}_{k-1 \text{ times}}, \overline{P}_{-1,2,\ldots,k}).$$

We claim that $g \in \mathcal{S}$. First, let us see that g is *strategy-proof*. Notice that agent 1 cannot manipulate g (because then he manipulates f as well). Assume that agent 2 manipulates g at (P_1, P_2) via P_2'. Then, $g(P_1, P_2') P_2 g(P_1, P_2)$ which implies

$$f(P_1, P_2', \ldots, P_2', \overline{P}_{-1,2,\ldots,k}) P_2 f(P_1, P_2, \ldots, P_2, \overline{P}_{-1,2,\ldots,k}). \tag{7}$$

By the *strategy-proofness* of f,

$$f(P_1, P_2, \ldots, P_2, \overline{P}_{-1,2,\ldots,k}) R_2 f(P_1, P_2', P_2, \ldots, P_2, \overline{P}_{-1,2,\ldots,k}) R_2 \ldots R_2 f(P_1, P_2', \ldots, P_2', \overline{P}_{-1,2,\ldots,k}).$$

By transitivity, $f(P_1, P_2, \ldots, P_2, \overline{P}_{-1,2,\ldots,k}) R_2 f(P_1, P_2', \ldots, P_2', \overline{P}_{-1,2,\ldots,k})$, contradicting (7). Thus, g is *strategy-proof*.

To see that g is *unanimous*, notice first that if $\check{P}_1^x, \check{P}_2^x \in \mathcal{P}$, then $g(\check{P}_1^x, \check{P}_2^x) = x$ by (6) and *strategy-proofness*. Next, consider any $w \in X \setminus \{x\}$. Let $P_i^\star \in \mathcal{P}$ be such that $P_i^\star : w, x, \ldots$ for each $i \in \{2, \ldots, k\}$. Then, by *strategy-proofness*, $f(P_1', P_2^\star, \overline{P}_{-1,2}) \in \{x, w\}$ (otherwise agent 2 manipulates f at $(P_1', P_2^\star, \overline{P}_{-1,2})$ via $\overline{P}_2$). Again by *strategy-proofness*, changing the preferences of agents 3 to k one at a time, we get $f(P_1', P_2^\star, \ldots, P_k^\star, \overline{P}_{-1,2,\ldots,k}) \in \{x, w\}$. Since $N_x(P_1', P_2^\star, \ldots, P_k^\star, \overline{P}_{-1,2,\ldots,k}) = \emptyset$, by Lemma 3, $f(P_1', P_2^\star, \ldots, P_k^\star, \overline{P}_{-1,2,\ldots,k}) = w$. Furthermore, let $\widetilde{P}_1 \in \mathcal{P}$ with $\widetilde{P}_1 : w, x, \ldots$ By *strategy-proofness*, $f(\widetilde{P}_1, P_2^\star, \ldots, P_k^\star, \overline{P}_{-1,2,\ldots,k}) \in \{x, w\}$ (otherwise agent 1 manipulates f at $(\widetilde{P}_1, P_2^\star, \ldots, P_k^\star, \overline{P}_{-1,2,\ldots,k})$ via P_1'). Again by Lemma 3, since $N_x(\widetilde{P}_1, P_2^\star, \ldots, P_k^\star, \overline{P}_{-1,2,\ldots,k}) = \emptyset$, we have

$$f(\widetilde{P}_1, P_2^\star, \ldots, P_k^\star, \overline{P}_{-1,2,\ldots,k}) = w. \tag{8}$$

Let $P_1^w, P_2^w \in \mathcal{P}$. Then $g(P_1^w, P_2^w) = w$ by *strategy-proofness* and (8). Thus, $g \in \mathcal{S}$ and, as we already proved for the $n = 2$ case, g is *own-top-only*.

To finish the proof, consider any $P_2^x \in \mathcal{P}$. By *strategy-proofness* and (6), $g(P_1', P_2^x) = x$. Since g is *own-top-only*, $g(P_1^\star, P_2^x) = x$. By the definition of g, $f(P_1^\star, P_2^x, \ldots, P_2^x, \overline{P}_{-1,2,\ldots,k}) = x$. Now, by *strategy-proofness*, we can change one at a time the preferences of agents 2 to k to obtain $f(P_1^\star, \overline{P}_{-1}) = x$. This contradicts (3). Therefore, f is *own-top-only* and, hence, *tops-only*. □

The proof of Lemma 4 follows closely the ideas in the proof of Theorem 2 (b) in Ninjbat (2012). Note, however, that Ninjbat (2012) does not analyze *tops-only* rules. By Lemmata 1 and 4 the next result follows.

Corollary 1 $\mathcal{S} \subseteq \mathcal{T}$.

4 Dual order structure of $\mathcal{T}$ and the proof

Before presenting the comparability criteria, we state a result analogous to Lemma 3 but now for rules in $\mathcal{T}$.

Lemma 5 *Let $f \in \mathcal{T}$, $P \in \mathcal{P}^n$ and $x \in X$ be such that $f(P) = x$. Then, $N_x(P) \neq \emptyset$.*

Proof. Let $f \in \mathcal{T}$ and assume there is $P \in \mathcal{P}^n$ such that $t(P_i) \neq f(P)$ for each $i \in N$. Let $P^\star \in \mathcal{P}^n$ be such that, for each $i \in N$, $t(P_i^\star) = t(P_i)$ and $yP_i^\star f(P)$ for each $y \in X \setminus \{f(P)\}$. By *tops-only*, $f(P^\star) = f(P)$. Let $x \in X \setminus \{f(P)\}$. Then, $xP_i^\star f(P^\star)$ for each $i \in N$, contradicting *efficiency*. Therefore, $N_x(P) \neq \emptyset$. □

Let $f \in \mathcal{T}$ and let $P \in \mathcal{P}^n$. Profile P is *manipulable for* f if there are $i \in N$ and $P_i^\star, P_i' \in \mathcal{P}$ such that $t(P_i^\star) = t(P_i)$ and i manipulates f at $(P_i^\star, P_{-i})$ via P_i'. Denote by M_f the set of all manipulable profiles for f. Our comparability criterion with respect to manipulability is presented next.

Definition 1 *Let $f, g \in \mathcal{T}$. We say that* ***f is at least as manipulable as g,*** *and write* $\boldsymbol{f \succeq_m g}$*, if $|M_f| \geq |M_g|$.*

It is clear that $f \in \mathcal{T}$ is *strategy-proof* if and only if $M_f = \emptyset$. Thus, the next remark follows.

Remark 1 Let $f \in \mathcal{T}$. Then, $f \in \mathcal{S}$ if and only if $g \succeq_m f$ for each $g \in \mathcal{T}$.

Let $f \in \mathcal{T}$ and let $P \in \mathcal{P}^n$. Profile P is *dictatorial for f* if, for each $i \in N$ such that $t(P_i) \neq f(P)$, we have $f(P_i', P_{-i}) = f(P)$ for each $P_i' \in \mathcal{P}$. Denote by D_f the set of all dictatorial profiles for f. Our comparability criterion with respect to dictatorial power is presented next.

Definition 2 *Let $f, g \in \mathcal{T}$. We say that* ***f is at least as dictatorial as g,*** *and write* $\boldsymbol{f \succeq_d g}$*, if $|D_f| \geq |D_g|$.*

The only rules for which all profiles are dictatorial are precisely *dictatorial* rules.

Lemma 6 *Let $f \in \mathcal{T}$. Then, $f \in \mathcal{D}$ if and only if $D_f = \mathcal{P}^n$.*

Proof. Let $f \in \mathcal{T}$. It is clear that if f is *dictatorial*, then $D_f = \mathcal{P}^n$. Let f be such that $D_f = \mathcal{P}^n$. We start with the case $n = 2$. If f is not *dictatorial* then for each $i \in \{1, 2\}$ there is $P^i \in \mathcal{P}^2$ such that $f(P^i) \neq t(P^i_i)$. Thus, by Lemma 5,

$$f(P^1) = t(P^1_2) \text{ and } f(P^2) = t(P^2_1). \tag{9}$$

Consider now profile (P^1_1, P^2_2). First, assume $t(P^1_1) \neq t(P^2_2)$. By Lemma 5, w.l.o.g., we have $f(P^1_1, P^2_2) = t(P^1_1)$. Since profile (P^1_1, P^2_2) is dictatorial and $f(P^1_1, P^2_2) \neq t(P^2_2)$, it follows that $f(P^1_1, P^1_2) = t(P^1_1)$, contradicting (9). Next, assume $t(P^1_1) = t(P^2_2)$. Let $\widetilde{P}_1 \in \mathcal{P}$ be such that $t(\widetilde{P}_1) \notin \{t(P^1_1), t(P^1_2)\}$. Since profile P^1 is dictatorial and $f(P^1) \neq t(P^1_1)$, it follows that $f(\widetilde{P}_1, P^1_2) = t(P^1_2)$. Then, as $t(\widetilde{P}_1) \neq t(P^2_2)$, the argument follows as before and we reach another contradiction. Hence, f is dictatorial when $n = 2$. Assume now that $n \geq 3$ and that every $n-1$ agent rule $f' \in \mathcal{T}$ such that $D_{f'} = \mathcal{P}^{n-1}$ is *dictatorial*. Given $f : \mathcal{P}^n \longrightarrow X$ define rule $g : \mathcal{P}^{n-1} \longrightarrow X$ as follows.[3] For each $(P_1, P_3, \ldots, P_n) \in \mathcal{P}^{n-1}$,

$$g(P_1, P_3, \ldots, P_n) = f(P_1, P_1, P_3, \ldots, P_n).$$

Clearly, g is *efficient*, *tops-only*, and $D_g = \mathcal{P}^{n-1}$. Therefore, g has a dictator, say $i^\star$. First, assume $i^\star \in \{3, \ldots, n\}$. We now show that $i^\star$ is also the dictator of f. If $i^\star$ is not the dictator

[3] The idea of defining a $n-1$ agent rule from an n agent rule by "coalescing" two agents is due to Sen (2001).

of f, then there is $(P_1, P_2, P_3, \dots, P_n) \in \mathcal{P}^n$ and $k \in N \setminus \{i^\star\}$ such that $f(P_1, P_2, P_3, \dots, P_n) = t(P_k) \neq t(P_{i^\star})$. Since f is *tops-only* and P is dictatorial, $f(P_k, P_2, P_3, \dots, P_n) = t(P_k)$. Again, since f is *tops-only* and P is dictatorial, $f(P_k, P_k, P_3, \dots, P_n) = t(P_k)$. By the definition of g, $g(P_k, P_3, \dots, P_n) = t(P_k)$, contradicting that $i^\star$ is the dictator of g. Therefore, $i^\star$ is the dictator of f as well. Next, assume $i^\star = 1$. Notice that $f(P_1, P_2, P_3, \dots, P_n) \in \{t(P_1), t(P_2)\}$ for each $(P_1, P_2, P_3, \dots, P_n) \in \mathcal{P}^n$ (otherwise, there is $j \in \{3, \dots, n\}$ such that $f(P_1, P_2, P_3, \dots, P_n) = t(P_j) \notin \{t(P_1), t(P_2)\}$ and, since $(P_1, P_2, P_3, \dots, P_n)$ is dictatorial, $f(P_1, P_1, P_3, \dots, P_n) = t(P_j)$, contradicting that 1 is the dictator of g). Let $(\overline{P}_3, \dots, \overline{P}_n) \in \mathcal{P}^{n-2}$. Define rule $h : \mathcal{P}^2 \longrightarrow X$ as follows. For each $(P_1, P_2) \in \mathcal{P}^2$,

$$h(P_1, P_2) = f(P_1, P_2, \overline{P}_3, \dots, \overline{P}_n).$$

Clearly, h is *efficient*, *tops-only*, and $D_h = \mathcal{P}^2$. Therefore, as we already proved for the case $n = 2$, h has a dictator. It remains to be shown that this dictator does not depend on the sub-profile $(\overline{P}_3, \dots, \overline{P}_n)$ chosen to define h. Assume it does. Then, there is another sub-profile $(\widetilde{P}_3, \dots, \widetilde{P}_n) \in \mathcal{P}^{n-1}$ such that, w.l.o.g.,

$$f(P_1, P_2, \overline{P}_3, \dots, \overline{P}_n) = t(P_1) \text{ and } f(P_1, P_2, \widetilde{P}_3, \dots, \widetilde{P}_n) = t(P_2), \tag{10}$$

and also $t(P_1) \neq t(P_2)$. Let $z \in X \setminus \{t(P_1), t(P_2)\}$ and consider any $P_3^z \in \mathcal{P}$. Then,

$$f(P_1, P_2, P_3^z, \overline{P}_{-1,2,3}) \in \{t(P_1), z\}$$

(otherwise, since $(P_1, P_2, P_3^z, \overline{P}_{-1,2,3})$ is dictatorial, $f(P_1, P_2, P_3^z, \overline{P}_{-1,2,3}) \notin \{t(P_1), z\}$ implies $f(P_1, P_2, \overline{P}_3, \dots, \overline{P}_n) \notin \{t(P_1), z\}$, a contradiction). Using the same argument, changing the preference of one agent at a time, it follows that $f(P_1, P_2, P^z_{-1,2}) \in \{t(P_1), z\}$. Similarly, but now starting from $f(P_1, P_2, \widetilde{P}_3, \dots, \widetilde{P}_n)$, we get $f(P_1, P_2, P^z_{-1,2}) \in \{t(P_2), z\}$. Thus, $f(P_1, P_2, P^z_{-1,2}) = z$ and, since profile $(P_1, P_2, P^z_{-1,2})$ is dictatorial, $f(P_1, P_1, P^z_{-1,2}) = z$. Then, $g(P_1, P^z_{-1,2}) = z$ and agent 1 is not the dictator of g, a contradiction. Therefore, either agent 1 or agent 2 is the dictator of f. $\square$

The next remark follows from Lemma 6.

Remark 2 Let $f \in \mathcal{T}$. Then, $f \in \mathcal{D}$ if and only if $f \succeq_d g$ for each $g \in \mathcal{T}$.

It turns out that the classification of a preference profile as manipulable or dictatorial, for a given rule, is exhaustive.

Lemma 7 *Let $f \in \mathcal{T}$ and let $P \in \mathcal{P}^n$. Then, P is either dictatorial for f or manipulable for f.*

Proof. Let $f \in \mathcal{T}$ and $P \in \mathcal{P}^n$. Assume P is not dictatorial for f. Then, there are $i \in N$ such that $t(P_i) \neq f(P)$ and $P_i' \in \mathcal{P}$ such that $f(P_i', P_{-i}) \neq f(P)$. Consider $P_i^\star \in \mathcal{P}$

such that $t(P_i^\star) = t(P_i)$ and $f(P_i', P_{-i}) P_i^\star f(P)$. By *tops-only*, $f(P_i', P_{-i}) = f(P)$. Thus, $f(P_i', P_{-i}) P_i^\star f(P_i^\star, P_{-i})$ and i manipulates f at $(P_i^\star, P_{-i})$ via P_i'. Therefore, P is manipulable for f. □

The dual order structure of $\mathcal{T}$ is an immediate consequence of Lemma 7.

Corollary 2 *Let $f, g \in \mathcal{T}$. Then, $f \succeq_d g$ if and only if $g \succeq_m f$.*

We are now finally in a position to prove the Gibbard–Satterthwaite Theorem.

Theorem 1 $\mathcal{S} = \mathcal{D}$.

Proof. It is clear that $\mathcal{D} \subseteq \mathcal{S}$. Next, we prove that $\mathcal{S} \subseteq \mathcal{D}$. Let $f \in \mathcal{S}$. By Corollary 1, $f \in \mathcal{T}$. By Remark 1, $g \succeq_m f$ for each $g \in \mathcal{T}$. By Corollary 2, $f \succeq_d g$ for each $g \in \mathcal{T}$. Thus, by Remark 2, $f \in \mathcal{D}$. □

References

ARRIBILLAGA, R. P. AND J. MASSÓ (2016): "Comparing generalized median voter schemes according to their manipulability," *Theoretical Economics*, 11, 547–586.

BARBERÀ, S. (2011): "Strategyproof social choice," *Handbook of Social Choice and Welfare*, 2, 731–831.

GIBBARD, A. (1973): "Manipulation of voting schemes: a general result," *Econometrica*, 41, 587–601.

MAUS, S., H. PETERS, AND T. STORCKEN (2007): "Anonymous voting and minimal manipulability," *Journal of Economic Theory*, 135, 533–544.

NINJBAT, U. (2012): "Another direct proof for the Gibbard–Satterthwaite Theorem," *Economics Letters*, 116, 418–421.

PATHAK, P. A. AND T. SÖNMEZ (2013): "School admissions reform in Chicago and England: Comparing mechanisms by their vulnerability to manipulation," *American Economic Review*, 103, 80–106.

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

SEN, A. (2001): "Another direct proof of the Gibbard–Satterthwaite theorem," *Economics Letters*, 70, 381–385.