BEYOND ARROW: REVEALING TRUE PREFERENCES

ARYAN ARORA

Abstract. In this paper, we examine the existence of strict strategy-proof voting mechanisms. Using Arrow's Impossibility Theorem we prove any rational social choice function that respects Pareto optimality, independence of irrelevant alternatives, and monotonicity must be dictatorial. Next, we use the Gibbard-Satterthwaite theorem to show that strategy-proof voting mechanisms are necessarily dictatorial. Then, we use the Gibbard-Satterthwaite correspondence theorem to establish a one-to-one correspondence between social choice functions under Arrow's construction and voting mechanisms under Gibbard and Satterthwaite's construction. Finally, we extend these results by considering other voting mechanisms and discussing Vickrey-Clarke-Groves mechanisms.

1. Introduction

The ancient Greeks had a direct rather than representative voting system. Any male over the age of twenty could vote and nearly one-fifth of Greeks did regularly [3]. Citizens voted by raising their hands or placing stones in buckets. But their democracy was far from perfect. Voting procedures often lead to loud and spirited arguments, much to the dismay of townspeople, and many could not vote either because they did not meet the eligibility requirements or because they could not be physically present for voting procedures. Today, the costs of voting have fallen but the challenges surrounding voting procedures remain. Nationally, there are strong critiques of our electoral college system [4]. Even smaller voting committees, like companies, academic departments, and thesis advising committees face voting challenges. These challenges include an inability to convert ordinal into cardinal rankings of alternatives, and doubts about whether individuals are casting votes that reveal their true preferences.

The second challenge is potentially more worrisome and the focus of this paper. This paper investigates the existence of strategy-proof voting mechanisms. A strategy-proof voting mechanism is a procedure for converting individual preferences into a societal choice in which an individual does not have an incentive to cast a vote that is different from their sincere preferences. Most simply, a strategy-proof voting mechanism induces every voter to reveal their true preferences.

Beginning with Arrow's impossibility theorem, the first section establishes that there is no non-dictatorial method for converting individual preferences into societal preferences. The second section alleviates the condition of a societal weak ordering and considers whether it is possible to identify a societal best alternative from individual preferences. In both cases, we show that it is not possible unless we adopt a dictatorial voting system. The end of the second section establishes a correspondence between Arrow's impossibility theorem and the Gibbard-Satterthwaite theorem. The final section considers extensions of these results. Throughout the paper, we aim to present the results proven by Arrow, Gibbard, and Satterthwaite and to supplement them with examples and mathematical intuition that make their results more accessible to audiences who do not work at the intersection of mathematics and microeconomic theory.

2. Arrows Impossibility Theorem

Arrow's impossibility theorem, from Kenneth Arrow's 1951 dissertation, proves that when two or more voters must choose between three or more distinct alternatives, no ranked electoral voting system can convert individual preferences into a societal ranking of preferences. Arrow's theorem can most simply be understood by considering three voters: v_1, v_2, and v_3, who are voting on three alternatives: α, β, and γ. Assume that the three individuals have the following preferences:

$$v_1 : \alpha > \beta > \gamma$$
$$v_2 : \beta > \gamma > \alpha$$
$$v_3 : \gamma > \alpha > \beta$$

Note that the majority of voters (voters v_1 and v_3) prefer alternative α to β. Note also that the majority of voters (voters v_1 and v_2) prefer alternative β to alternative γ. Then our societal ranking of alternatives should take the following form:

$$\alpha > \beta > \gamma$$

By transitivity, α is preferred to γ. But the majority of voters (voters v_2 and v_3) prefer alternative γ to alternative α. Here lies Arrow's impossibility theorem: there is no way to construct an appropriate societal ranking of alternatives. Although we have considered a reductive example, the numbers can be scaled to any finite value and our result holds. Consider now a society with 99 individuals such that 33 have identical preferences to v_1, 33 have identical preferences to v_2, and 33 have identical preferences to v_3. Observe that we encounter the same challenge in a voting committee of size 99 as we do in a voting committee of size three.

The same process can completed to construct an inductive chain on the number of alternatives. Consider putting any number of additional alternatives anywhere in the above preferences and note that if you follow the procedure we did above, the social preference ranking will still have alternative α preferred to alternative γ, despite the majority of voters preferring the converse. One nice way of expressing the key idea of Arrow's impossibility theorem is that it is difficult to construct well-behaved preferences of groups, even if individuals have well-behaved preferences.

2.1. A Proof of Arrows Impossibility Theorem [1, 6].

Let $A = \{\alpha, \beta, \gamma, ..., \omega\}$ be a finite set of alternatives. We will denote an individual's transitive preference of the set of alternatives with ties allowed as $T_{|A|}$. Because we allow ties we call their preferences a weak ordering over the alternatives. If, instead, we did not allow ties the individuals would have strict orderings over the alternatives.

Consider a society with n individuals who each carry potentially distinct transitive preferences over A. For example, consider individuals who must choose between the alternatives $A = \{\alpha, \beta, \gamma\}$. The set of all transitive preferences is:

$$\begin{aligned} T_{|A|} = \{&(\alpha > \beta > \gamma), (\alpha \geq \beta > \gamma), (\alpha > \beta \geq \gamma), (\alpha \geq \beta \geq \gamma), \\ &(\alpha > \gamma > \beta), (\alpha \geq \gamma > \beta), (\alpha > \gamma \geq \beta), (\alpha \geq \gamma \geq \beta), \\ &(\beta > \gamma > \alpha), (\beta \geq \gamma > \alpha), (\beta > \gamma \geq \alpha), (\beta \geq \gamma \geq \alpha), \\ &(\beta > \alpha > \gamma), (\beta \geq \alpha > \gamma), (\beta > \alpha \geq \gamma), (\beta \geq \alpha \geq \gamma), \\ &(\gamma > \alpha > \beta), (\gamma \geq \alpha > \beta), (\gamma > \alpha \geq \beta), (\gamma \geq \alpha \geq \beta), \\ &(\gamma > \beta > \alpha), (\gamma \geq \beta > \alpha), (\gamma > \beta \geq \alpha), (\gamma \geq \beta \geq \alpha)\} \end{aligned}$$

A *constitution*, $C : T^n_{|A|} \to S_{|A|}$, is a function that aggregates individual transitive preferences into a *social choice*: a societal weak ordering of alternatives. $T^n_{|A|}$ is all possible preferences for all

members of society and $S_{|A|} \in T_{|A|}$. For example, a particular constitution, C_i, might map the following preferences for three voters:

$$T_{3,1} : \alpha > \beta > \gamma$$
$$T_{3,2} : \beta > \gamma > \alpha$$
$$T_{3,3} : \gamma > \alpha > \beta$$

to the following social preference: $S_3 = (\alpha > \beta > \gamma)$. We seek for our constitution to satisfy the following properties:

Definition 1 (Pareto Optimality)**.** *If every individual in a society prefers alternative α to β, our constitution should produce a social preference that ranks α higher than β.*

Assume that we have a committee with five individuals who are choosing between three alternatives, $A = \{\alpha, \beta, \gamma\}$. Additionally, assume the five individuals have the following preferences:

$$\begin{aligned} T_{3,1} &: \quad \alpha > \beta > \gamma \\ T_{3,2} &: \quad \alpha > \gamma > \beta \\ T_{3,3} &: \quad \beta > \alpha > \gamma \\ T_{3,4} &: \quad \alpha > \gamma > \beta \\ T_{3,5} &: \quad \alpha > \beta > \gamma \end{aligned}$$

Note that although all members do not have identical preferences, all voters prefer alternative α to alternative γ. If the constitution obeys Pareto optimality, the social choice must also have alternative α ranked before alternative γ.

Definition 2 (Independence of Irrelevant Alternatives)**.** *Assume our constitution produces a societal ranking of alternatives such that α is preferred to β and β is preferred to γ. Now assume that individuals change their preferences so that γ is preferred to β. In the societal ranking, α must still be preferred to β.*

Assume, as we previously did, that we are considering a voting committee with 5 individuals who each carry transitive preferences over an alternative set $A = \{\alpha, \beta, \text{ and } \gamma\}$ which we denote $T_{|A|,i}$ as shown below. Assume that all individuals change their preferences to $T'_{|A|,i}$ shown below.

$$\begin{aligned} T_{3,1} &: \alpha > \beta > \gamma \\ T_{3,2} &: \alpha > \beta > \gamma \\ T_{3,3} &: \alpha > \beta > \gamma \\ T_{3,4} &: \alpha > \beta > \gamma \\ T_{3,5} &: \alpha > \beta > \gamma \end{aligned} \qquad \begin{aligned} T'_{3,1} &: \alpha > \gamma > \beta \\ T'_{3,2} &: \alpha > \gamma > \beta \\ T'_{3,3} &: \alpha > \gamma > \beta \\ T'_{3,4} &: \alpha > \gamma > \beta \\ T'_{3,5} &: \alpha > \gamma > \beta \end{aligned}$$

Because $\alpha > \beta$ for all $T_{|A|,i}$, $T'_{|A|,i}$ the societal ranking should still rank α before β. The relative societal ranking of α and β should not be impacted by another alternative: γ.

Definition 3 (Nondictatorship)**.** *No individual alone determines the relative ranking of two alternatives for society.*

Theorem 1 (*Arrow's Impossibility Theorem*)**.** *If $|A| \geq 3$ and $2 < n < \infty$, any social choice function that respects transitivity, Pareto optimality, and independence of irrelevant alternatives must violate non-dictatorship.*

We will begin by mathematically formalizing the ideas we have discussed. Let $|A| = m, m \in \mathbb{N}$ and $m \geq 3$. Additionally, let $V = \{v_1, v_2, ..., v_n\}$ be the set of voters such that that $|V| = n$, $n \in \mathbb{N}$. Let $p : V \times A \to \{1, 2, ..., m\}$ be a preference function that assigns each voter a ranking of the elements of A such that $p(v_i, \alpha)$ is the ranking of alternative α given by voter v_i. It follows that voter v_i ranks α higher than β if and only if $p(v_i, \alpha) > p(v_i, \beta)$. Let a preference state, p_i be a set containing sets of n votes: one from each member of society. For example, $p_1 = \{v_1, v_2, ..., v_n\}$, $p_2 = \{v'_1, v'_2, ..., v'_n\}$, $p_i = \{v^*_1, v^*_2, ..., v^*_n\}$ where $v, v', v^* \in T_{|A|}$ are all different votes that an individual can cast. Additionally, let $P = \{p_1, p_2, ..., p_i, ...\}$ be the set of all possible preference states. Then $|P| = (m! \cdot 2^{m-1})^n$. In a society with n individuals and m alternatives, each individual has $m! \cdot 2^{m-1}$ because there are $m!$ orderings of the alternatives and 2^{m-1} $m - 1$ inequalities which can be either strong, $>$, or weak, $\geq$. In an n member society, there are $(m! \cdot 2^{m-1})^n$ different preference states.

Let $C : P \times A \to \{1, 2, ..., m\}$ be the constitution which takes in both the preference state and a particular candidate and outputs a rank. Thus $C(p_i, \alpha) > C(p_i, \beta)$ indicates that given a particular preference state, p_i, society prefers alternative α to β. Finally let $R(p, \alpha, \beta)$ for $p \in P, \alpha, \beta \in A$ be the set of voters in preference state p who rank $\alpha > \beta$. Given our framework, we can mathematically formalize our previous definitions.

Definition 4 (Pareto Optimality)**.** *Let $\alpha, \beta \in A$ and let $\alpha > \beta$ denote that α is preferred to β. Let the input set of our constitution be $V = (v_1, v_2, ..., v_n)$. If $\alpha > \beta$ for all $v_i \in V$, $\alpha > \beta$ in $S_{|A|}$.*

Definition 5 (Independence of Irrelevant Alternatives)**.** *Let $\alpha, \beta, \gamma \in A$ and let $\alpha > \beta$ and $\beta > \gamma$ in $S_{|A|}$. Now assume that voters change their preferences so that $\gamma > \beta$ in $S_{|A|}$. It must still hold that $\alpha > \beta$ in $S_{|A|}$.*

We will now prove that Pareto optimality and independence of irrelevant alternatives conditions produce another condition called *monotonicity.*

Definition 6 (Monotonicity)**.** *Assume we have two distinct preference states: p_a and p_b. Additionally assume that under p_a, our constitution ranks alternative α higher than alternative β. If the only difference between p_a and p_b is that some of the voters who rank $\beta > \alpha$ in p_a rank $\alpha > \beta$ under preference state p_b, our constitution must still rank $\alpha > \beta$.*

Lemma 1. *If $C(p_m, \alpha) > C(p_m, \beta)$ and $R(p_n, \alpha, \beta) \subseteq R(p_m, \alpha, \beta)$ for some $p_m, p_n \in P$ and $\alpha, \beta \in A$, then $C(p_n, \alpha) > C(p_n, \beta)$.*

This lemma states that if the constitution for a preference state, p_m ranks $\alpha > \beta$ and the set of voters who rank α higher than β in p_m is a subset of the voters who rank α higher than β in preference state p_n, then the constitution for p_n ranks α higher than β

Proof. Let $p_m, p_n \in P$, $\alpha, \beta \in A$. Additionally, let $R(p_m, \alpha, \beta) \subseteq R(p_n, \alpha, \beta)$. We want to show that $C(p_n, \alpha) > C(p_n, \beta)$. Let $p^* \in P$ be the preference state that has the following qualities for voter $v_i \in V$ and alternative $\gamma \in A$ such that $\gamma \neq \alpha, \beta$:

$$\begin{gathered} p^*(v_i, \alpha) = p_n(v_i, \alpha) \\ p^*(v_i, \beta) = p_n(v_i, \beta) \\ p^*(v_i, \gamma) = p_m(v_i, \gamma) \end{gathered}$$

It follows that

$$\begin{gathered} C(p_m, \alpha) > C(p_m, \beta) \\ C(p^*, \gamma) = C(p_m, \gamma) \end{gathered}$$

Because $p^* = p_n$ for α and β and $R(p_m, \alpha, \beta) \subseteq R(p_n, \alpha, \beta)$, $p(v_i, \alpha) > p(v_i, \beta)$ implies $p^*(v_i, \alpha) > p^*(v_i, \beta)$. Then, $C(p^*, \alpha) > C(p^*, \beta)$. Finally, by independence of irrelevant alternatives, $C(p_n, \alpha) > C(p_n, \beta)$, as desired. □

Now that we have completed the proof, we will formalize our definitions of monotonicity and non-dictatorship and introduce the definition of an oligarchy.

Definition 7 (Monotonicity)**.** *Let p_a and p_b be two preference states such that $p_a \neq p_b$. If $C(p_a, \alpha) > C(p_a, \beta)$. Additionally let $C(p_a, \gamma) = C(p_b, \gamma)$ for all $\gamma \in A$ such that $\gamma \neq \alpha, \beta$. If $p_a(v_i, \alpha) > p_a(v_i, \beta)$ implies that $p_b(v_i, \alpha) > p_b(v_i, \beta)$, then $C(p_a, \alpha) > C(p_a, \beta)$.*

Definition 8 (Nondictatorship)**.** *There does not exist an $i \in \{1, 2, ..., n\}$ such that $p_k(v_i, \alpha) > p_k(v_i, \beta)$ implies $C(p_k, \alpha) > C(p_k, \alpha)$ for all $p_k \in P, \alpha, \beta \in A$.*

We now assume that C is a monotonic constitution that satisfies both Pareto optimality and independence of irrelevant alternatives. For all pairs of alternatives, α, β, let $R(p, \alpha, \beta) = \{v_i \in V \mid p(v_i, \alpha) > p(v_i, \alpha)\}$ or the set of voters who prefer α to β under a particular preference state, p.

Definition 9 (Oligarchy)**.** *A set, $X \subset V$, is an oligarchy for an alternative α over an alternative β if for all preference states $p \in P$, $X \subseteq R(p, \alpha, \beta)$ implies that $p(v, \alpha) > p(v, \beta)$.*

More simply, a set of voters is an oligarchy if the relative preferences of that group between two alternatives determine the relative societal preferences over those alternatives. Let $\mathcal{U}\{\alpha, \beta\}$ denote the set of oligarchies for an alternative α over an alternative β and let $\mathcal{U}$ be the set of all oligarchies. We will proceed by showing that $\mathcal{U}$ is an *ultrafilter*.

Definition 10 (Ultrafilter)**.** *Let A be a set and let $\mathcal{U} \subseteq \mathcal{P}(A)$, the power set of A. $\mathcal{U}$ is an* ultrafilter *if:*

- *$\mathcal{U}$ is **upward closed**: for all $C \in \mathcal{P}(A)$, for all $B \in \mathcal{U}$, $B \subseteq C$ implies $C \subseteq \mathcal{U}$.*
- *$\mathcal{U}$ is **closed under finite intersections**: for all $B_i \in \mathcal{U}$, $B_a \cap B_b \cap ... \cap B_n \in \mathcal{U}$*
- *$\mathcal{U}$ is **maximal**: For all $B \subseteq A$, exactly one of B and $A - B \in \mathcal{U}$.*

Lemma 2. *If $C(p, \alpha) > C(p, \beta)$ for some $p \in P$, $\alpha, \beta \in A$ and $X = R(p, \alpha, \beta)$, then $X = \mathcal{U}\{\alpha, \beta\}$.*

This lemma states that if the constitution ranks $\alpha > \beta$ for some preference state $p \in P$ and X is the set of voters who rank α higher than β in preference state p, X is an oligarchy for α over β.

Proof. In lemma 1 we showed that if $C(p_m, \alpha) > C(p_m, \beta)$ and $R(p_m, \alpha, \beta) \subseteq R(p_m, \alpha, \beta)$ for some $p_m, p_n \in P$ and $\alpha, \beta \in A$, then $C(p_n, \alpha) > C(p_n, \beta)$. We know that in preference state p, $X \in R(p, \alpha, \beta)$ implies that $\alpha > \beta$. Then, by the previous lemma, for any other preference state, if $X \in R(p, \alpha, \beta)$, the constitution will rank $\alpha > \beta$. Therefore, for all preference states $p' \in P$, $X \in R(p', \alpha, \beta)$ implies the constitution will rank $\alpha > \beta$. By definition of oligarchy, X is an oligarchy. □

Lemma 3. *If X is an oligarchy for α over β, then X is an oligarchy.*

Proof. We want to prove that $\mathcal{U}\{\alpha, \beta\} = \mathcal{U}$. We will prove this by showing that for all $\alpha \neq \beta$ and $\epsilon \neq \zeta$, $\mathcal{U}\{\alpha, \beta\} = \mathcal{U}\{\epsilon, \zeta\}$ or the same group of voters who constitute an oligarchy for α over β constitute an oligarchy for ϵ over ζ for all $\alpha, \beta, \epsilon, \zeta \in A$.

Fix $\alpha, \beta \in A$ and choose ϵ such that $\epsilon \neq \beta$. Let $X \in \mathcal{U}\{\alpha, \beta\}$ and $X \subseteq R(p, \alpha, \beta)$. Choose a preference state p_ϵ such that $R(p, \alpha, \beta) = R(p, \epsilon, \beta)$. We know that such a set exists because it can be constructed by taking all voters in $R(p, \epsilon, \beta)$ and flipping the rankings of α and ϵ for all $v_i \in p$. Because $X \in \mathcal{U}\{\alpha, \beta\}$ it follows that $C(p, \alpha) > C(p, \beta)$. Then, for p_ϵ we must have $C(p_\epsilon, \epsilon) > C(p_\epsilon, \beta)$. It follows that that $X \in \mathcal{U}\{\epsilon, \beta\}$. So $X \in \mathcal{U}\{\alpha, \beta\}$ implies $X \in \mathcal{U}\{\epsilon, \beta\}$. Thus

we have shown that the set of voters who compose an oligarchy are independent of the choice of alternatives, as desired. □

The result from the previous lemma showed that if $X = \mathcal{U}\{\alpha, \beta\}$, then $X = \mathcal{U}\{\epsilon, \zeta\}$ for all $\alpha, \beta, \epsilon, \zeta \in A$ because by replacing α with ϵ and β with ζ in the preferences where $X = \mathcal{U}\{\alpha, \beta\}$ we get $X = \mathcal{U}\{\alpha, \beta\}$. This means that an oligarchy for one pair of alternatives is an oligarchy over all alternatives.

Lemma 4. *$\mathcal{U}$ is an ultrafilter.*

Proof. Recall, by definition of ultrafilter, that to show that $\mathcal{U}$ is an ultrafilter we must show that $\mathcal{U}$ is upward closed, closed under finite intersections, and maximal.

We will first prove that $\mathcal{U}$ is upward closed by showing that $X \subseteq \mathcal{U}\{\alpha, \beta\}$ implies $Y \subseteq \mathcal{U}\{\alpha, \beta\}$ for all Y such that $X \subseteq Y$. Let $X \subseteq \mathcal{U}\{\alpha, \beta\}$ and let $Y \subseteq V$ such that $X \subseteq Y \subseteq V$. Additionally, let $p \in P$ be an arbitrary preference state such that $Y \subseteq R(p, \alpha, \beta)$. Because $X \subseteq Y \subseteq R(p, \alpha, \beta)$, $X \subseteq R(p, \alpha, \beta)$. Recall, by definition of oligarchy that $X \subseteq R(p, \alpha, \beta)$ implies $C(p, \alpha) > C(p, \beta)$. Therefore, $Y \subseteq \mathcal{U}\{\alpha, \beta\}$, as desired.

Next, we will show that $\mathcal{U}$ is closed under finite intersections. Let $X, Y \in \mathcal{U}\{\alpha, \beta\}$. Additionally, let p be a preference state and $\alpha, \beta \in A$ two alternatives such that $X \cap Y \subseteq R(p, \alpha, \beta)$. By assumption, $|A| \geq 3$ so there exists $\gamma \in A$ such that $\gamma \neq \alpha, \beta$. Suppose $p' \in P$ is a preference state such that:

(A) $p'(v_i, \alpha) > p'(v_i, \gamma)$ for all v_i such that $v_i \in X$ and $v_i \notin Y$
(B) $p'(v_i, \alpha) > p'(v_i, \gamma) > p'(v_i, \beta)$ for all $v_i \in X \cap Y$
(C) $p'(v_i, \gamma) > p'(v_i, \beta)$ for all v_i such that $v_i \in Y$ and $v_i \notin X$
(D) $R(p', \alpha, \beta) = R(p, \alpha, \beta)$

Recall that $X \in \mathcal{U}$ implies $C(p', \alpha) > C(p', \gamma)$. Then, $Y \in \mathcal{U}$ implies $C(p', \gamma) > C(p', \beta)$. Therefore:

$$C(p', \alpha) > C(p', \gamma) > C(p', \beta)$$

By independence of irrelevant alternatives, $C(p, \alpha) > C(p, \beta)$. But $X \in \mathcal{U}\{\alpha, \beta\}$ and $Y \in \mathcal{U}\{\alpha, \beta\}$, therefore $X \cap Y \in \mathcal{U}\{\alpha, \beta\}$. So for all $X, Y \in \mathcal{U}$, $X \cap Y \in \mathcal{U}$. Therefore, $\mathcal{U}$ is closed under finite intersections, as desired.

Finally, we must show that $\mathcal{U}$ is maximal. Recall, by definition of maximal, we must show that for all $X \subseteq V$, either X or all v_i such that $v_i \in V$ and $v_i \notin X \in \mathcal{U}\{\alpha, \beta\}$. Let $X \subseteq V$ and let p be a preference state such that for all $v_i \in X$, $\alpha > \beta$. If $C(p, \alpha) > C(p, \beta)$, then $X \subseteq \mathcal{U}\{\alpha, \beta\}$. If $C(p, \alpha) > C(p, \beta)$ whenever $X \subseteq R(p, \alpha, \beta)$, then $C(p', \alpha) > C(p', \beta)$ for $p' \in P$ in which $X = R(p', \alpha, \beta)$ and all v_i such that $v_i \in V$ and $v_i \notin X = R(p', \beta, \alpha)$. The previous statement asserts that the constitution will rank $\alpha > \beta$ for any preference state such that all voters in X rank $\alpha > \beta$ and all voters not in X rank $\beta > \alpha$. Then $\{v_i \,|\, v_i \in V, v_i \notin X\} \neq \mathcal{U}\{\beta, \alpha\}$. We have previously shown that $\mathcal{U}\{\beta, \alpha\} = \mathcal{U}\{\alpha, \beta\}$. Then $\{v_i \mid v_i \in V, v_i \notin X\} \neq \mathcal{U}\{\alpha, \beta\}$

Let $p \in P$ remain the preference state where $\alpha > \beta$ for all $v_i \in X$. Towards contradiction, assume $C(p, \beta) > C(p, \alpha)$. It follows that there exists some subset $Y \subseteq \{v_i \mid v_i \in V, v_i \notin X\}$ such that $p(v_i, \beta) > p(v_i, \alpha)$ for all $v_i \in Y$ and $Y \in \mathcal{U}\{\beta, \alpha\}$. But, by upward closure this implies that $\{v_i \mid v_i \in V, v_i \notin X\} \in \mathcal{U}\{\beta, \alpha\}$ which implies $\{v_i \mid v_i \in V, v_i \notin X\} \in \mathcal{U}\{\alpha, \beta\}$.

Thus $X \in \mathcal{U}\{\alpha, \beta\}$ necessarily implies $\{v_i \mid v_i \in V, v_i \notin X\} \notin \mathcal{U}\{\alpha, \beta\}$. This proves that from every set and its complement, only one can be an oligarchy. Because we know that X is an oligarchy and we know all Y such that $X \subseteq Y$ is an oligarchy, we know that none of their complements can be an oligarchy. Therefore X is maximal.

Then we have shown that $\mathcal{U}\{\alpha, \beta\}$ is an ultrafilter. Because $\mathcal{U}\{\alpha, \beta\} = \mathcal{U}$ we have consequently shown that $\mathcal{U}$ is an ultrafilter. □

The first result shows that, because X is an oligarchy, any set that contains X must also be an oligarchy. The voters from set X, by definition of being an oligarchy, choose the rankings for society. Adding more voters does not change this property. Then any set of voters, Y, such that $X \subseteq Y$ has all the necessary voters (from X) to constitute an oligarchy. The second result shows that, if X is an oligarchy and Y is an oligarchy, their intersection must also constitute a set of voters who are an oligarchy. Consider if this was not the case. Then there must be some voters in one set, but not in the other that constitute an oligarchy. But, in Lemma 3 we proved that this cannot be the case. The final result showed that an oligarchy must contain all elements of $\mathcal{U}$. We have now shown that $\mathcal{U}$ is an ultrafilter. The next lemma establishes a *finitely additive 0-1 measure* on V.

Definition 11 (Finitely Additive 0-1 Measure)**.** *Let V be a set. A finitely additive 0-1 measure is an injective function μ from the power set $\mathcal{P}(V)$ to $\{0,1\}$ which obeys the following constraints:*

(A) For all $J \subseteq V$, $\mu(J) \in \{0,1\}$
(B) $\mu(\emptyset) = 0$
(C) $\mu(V) = 1$
(D) For all $J_1, J_2, ..., J_n \in V$, if $J_1 \bigcap J_2 \bigcap ... \bigcap J_n = \emptyset$, then $\mu(J_1 \bigcup J_2 \bigcup ... \bigcup J_n) = \mu(J_1) + \mu(J_2) + ... + \mu(J_n)$

Define a function, $\mathcal{F}$ which maps $J \subseteq V$ to $\{0,1\}$ as follows:

$$\mathcal{F}(J) = \begin{cases} 1 \text{ if } \mathcal{U} \in J \\ 0 \text{ otherwise} \end{cases}$$

Lemma 5. *$\mathcal{F}(J)$ is a finitely additive 0-1 measure.*

Proof. The function $\mathcal{F}$ maps sets of voters to $\{0,1\}$. Note that each set of voters either contains $\mathcal{U}$, the set of voters who constitute an oligarchy, or does not. Then each set of voters is either mapped to 0 or 1, but never both. It follows that $\mathcal{F}$ is injective. Additionally, $\emptyset$ is the set with no voters. If $\mathcal{U} \neq \emptyset$ our constitution is not Pareto optimal. So $\mathcal{U} \neq \emptyset$. It follows that $\mu(\emptyset) = 0$. In the last lemma, we showed $\mathcal{U} \in V$. Then $\mu(V) = 1$. Finally, let $J_1, J_2, ..., J_n \in V$ be arbitrary. Then there are two cases:

(A) $\mathcal{U} \notin J_1, J_2, ..., J_n$
(B) $\mathcal{U} \in J_k$ and $\mathcal{U} \notin J_1, ..., J_{k-1}, J_{k+1}, ..., J_n$ for some $k \in \{1, 2, ..., n\}$

In the first case $\mathcal{U} \notin J_1, J_2, ..., J_n$. Therefore $\mathcal{U} \notin J_1 \bigcup J_2 \bigcup ... \bigcup J_n$ and $\mathcal{F}(J_1) = \mathcal{F}(J_2) = ... = \mathcal{F}(J_n) = 0$. Then $\mathcal{F}(J_1) + \mathcal{F}(J_2) + ... + \mathcal{F}(J_n) = \mathcal{F}(J_1 \bigcup J_2 \bigcup ... \bigcup J_n) = 0$. In the second case $\mathcal{U} \in J_1 \bigcup ... \bigcup J_k \bigcup ... \bigcup J_n$. Then $\mathcal{F}(J_k) = \mathcal{F}(J_1 \bigcup ... \bigcup J_k \bigcup ... \bigcup J_n) = 1$ and $\mathcal{F}(J_l) = 0$ for all $l \neq k$. So $\mathcal{F}(J_1) + ... + \mathcal{F}(J_k) + ... + \mathcal{F}(J_n) = 1 = \mathcal{F}(J_1 \bigcup ... \bigcup J_k \bigcup ... \bigcup J_n)$. Thus we have shown that $\mathcal{F}$ is a finitely additive 0-1 measure. □

We now introduce the notion of a principal measure to establish a mathematical structure which represents a dictator and then prove that our set of voters must have a principal measure.

Definition 12 (Principal Measure)**.** *A finitely additive 0-1 measure, $\mathcal{L}$, is principal if there exists $v_d \in V$ such that for all $J \subseteq V$, $\mathcal{L}(J) = 1$ if and only if $v_d \in J$.*

Lemma 6. *If V is finite and $\mathcal{L}$ is a finitely additive 0-1 measure on V, $\mathcal{F}$ is a principal measure.*

Proof. Let $\mathcal{L}$ be a finitely additive 0-1 measure on V where V has n elements, $n \in \mathbb{N}$. Because V is finite, it is equal to the disjoint union of its singleton subsets: $\{v_1\}, \{v_2\}, ..., \{v_n\}$ where $\{v_i\} \in V$ for $1 \leq i \leq n$ and all v_i are unique. By finite additivity, it must be true that

$$\mathcal{L}(\{v_1\}) + \mathcal{F}(\{v_2\}) + ... + \mathcal{F}(\{v_n\}) = 1.$$

Additionally, by definition of $\mathcal{L}$ it must be true that $\mathcal{L}(\{v_i\}) = 0$ or $\mathcal{L}(\{v_i\}) = 1$ for all $v_i \in V$. Because their sum is 1, it must be true that exactly one $v_i \in V$ has the property $\mathcal{L}(\{v_i\}) = 1$. Then, by definition of principal measure $\mathcal{L}$ is a principal measure. □

By construction, V is finite and we have already shown that $\mathcal{F}$ is a finitely additive 0-1 measure on V. Then, by our previous lemma, $\mathcal{F}$ is a principal measure and our voting committee has a dictator. Thus we have shown for two or more individuals choosing between three or more alternatives, any social choice function that respects transitivity, Pareto optimality, and independence of irrelevant alternatives must violate non-dictatorship. This completes the proof of Arrow's impossibility theorem. ■

3. Strategy Proof Voting Mechanisms

In a 2023 Gallup poll, over 40% of Americans identified as politically independent. [5] Yet independent parties, in aggregate, have never earned more than a few percentage points in presidential elections. [4] Critics point to the Electoral College as the source of this paradox. But what is it about our electoral college that leads independent candidates to earn so little of the vote? This section uses the Gibbard-Satterthwaite theorem to investigate this question. The answer is that the Electoral College is not a strategy-proof voting mechanism and individuals often have incentives to misrepresent their preferences.

We will later formalize, using set theory, what constitutes a strategy-proof voting mechanism. For now accept the non-mathematical definition: a strategy-proof voting mechanism is a way of counting votes so that no individual has an incentive to misrepresent their preferences. To illustrate the point about the Electoral College being non-strategy-proof consider an individual who has the following preferences over three candidates for US president, an independent candidate (I), a democratic candidate (D), and a republican candidate (R):

$$I > D > R$$

Knowing that the independent candidate does not have a strong chance of winning the election, this voter is faced with a dilemma. They can vote for candidate I and hope that the independent candidate wins. But the democratic and republican candidates are in a tightly contested race and not voting for the democratic candidate could create a path for the republican candidate to win—the voter's least preferred alternative. Or they can vote for the democratic candidate, D. While this would not help them achieve their most desired outcome, it can help prevent their least preferred outcome. Voters often choose the latter strategy. But what does the structure of the Electoral College have to do with that? Because individual votes are aggregated into a single vote from their member of the electoral college, individuals find themselves in a dilemma between voting according to their preferences and 'throwing their vote away' or misrepresenting their preferences in hopes of improving the outcome of the election according to their preferences.

Now consider, in line with Arrow's impossibility theorem and the Gibbbard-Satterthwaite theorem, that individuals cast a weak ordering over the alternatives instead of casting a single vote for one candidate. Would this change the voter's strategy? It would not. If they employ a *sincere strategy* they would vote according to their preferences: $I > D > R$. If they employ a *sophisticated strategy* they would vote $D > I > R$. This is not a strategy-proof voting mechanism because an individual can employ a sophisticated strategy to guarantee a better outcome as measured by their sincere preferences. Voting dilemmas like these are not limited to political elections. They occur at

the national level, but also in boardrooms of companies, academic administrations, and even voting mechanisms among friends about where to eat dinner.

The Gibbard-Satterthwaite Theorem allows us to consider the strategy proofness condition of voting mechanisms using a structure very similar to Arrow's impossibility theorem. They show, in line with Arrow's result, that strategy-proof voting mechanisms are necessarily dictatorial. Additionally, they establish a correspondence between the strategy proofness condition and the conditions for Arrow's impossibility theorem.

3.1. **Proof of Gibbard-Satterthwaite Theorem [7].** The Gibbard-Satterthwaite Theorem, while quite similar to Arrow's impossibility theorem, differs in a few noteworthy ways. Under Arrow's construction, individual weak orderings of the alternatives were mapped to a societal weak ordering. Gibbard and Satterthwaite alleviate this restriction by only requiring for individual preferences to be mapped to a single alternative which we call the *committee choice*. One might think of this as the utility-maximizing choice for society. The mathematical construction that we used in the proof of Arrow's Impossibility Theorem used a preference function p which mapped individual votes and alternatives to a rank. Gibbard and Satterthwaite omit this mechanism in their proof because they are not concerned with the relative ranking of all alternatives. A similar mechanism reappears in the correspondence theorem. Finally, we will be starting fresh with new notation in this section. All notation used in the proof is defined below.

We begin by developing a structure very similar to our setup of Arrow's impossibility theorem. Let $S_m = \{x, y, z, ...\}$ be a finite set of alternatives with cardinality m. Additionally let $I_n = \{i_1, ..., i_n\}$ be a committee of n voters. Each $i \in I_n$ has weak order preferences, R_i, over the alternatives in S_m. All R_i are complete, reflexive, and transitive. Each $i \in I_n$ casts a ballot, B_i which is a weak order on S_m. Then we will let π_m represent the collection of possible ballots for one individual and π_m^n the set of possible ballots for all voting individuals. Consider the case where $S_3 = \{x, y, z\}$ and $n = 5$. Then

$$\begin{aligned}\pi_3 = \{&(x > y > z), (x \geq y > z), (x > y \geq z), (x \geq y \geq z),\\ &(x > z > y), (x \geq z > y), (x > z \geq y), (x \geq z \geq y),\\ &(y > z > x), (y \geq z > x), (y > z \geq x), (y \geq z \geq x),\\ &(y > x > z), (y \geq x > z), (y > x \geq z), (y \geq x \geq z),\\ &(z > x > y), (z \geq x > y), (z > x \geq y), (z \geq x \geq y),\\ &(z > y > x), (z \geq y > x), (z > y \geq x), (z \geq y \geq x)\}\end{aligned}$$

and

$$\begin{aligned}\pi_3^5 = \{&\{B_1, B_2, B_3, B_4, B_5\},\\ &\{B_1^*, B_2^*, B_3^*, B_4^*, B_5^*\},\\ &\{B_1', B_2', B_3', B_4', B_5'\}, ...\}\end{aligned}$$

where all B_i, B_i', B_i^*, etc. $\in \pi_3$

The ballots are counted by a voting mechanism, v^{nm}, which maps $B = (B_1, B_2, B_3, ..., B_n) \in \pi_m^n$ to the *committee choice*, a single alternative, $x \in S_m$. Then all v^{nm} have a range of either S_m or some nonempty subset of S_m. The size depends on whether there are alternatives which are preferred to all others but which individuals are indifferent between. For example, if everyone has

preferences $R_i = (x \geq y > z)$ the voting mechanism should produce both x and y because individuals have not indicated a clear preference between the two. To visualize a voting mechanism consider $n = 5, m = 3, S_m = \{x, y, z\}, B = (B_1, ..., B_5)$ and $B_i = (x > y > z)$ for all $i \in I_n$. Then

$$v^{5,3}(B) = v^{5,3} \begin{pmatrix} B_1: & x > y > z \\ B_2: & x > y > z \\ B_3: & x > y > z \\ B_4: & x > y > z \\ B_5: & x > y > z \end{pmatrix} = (x) \in S_m$$

Note that the voting mechanism we have described is very similar to the social choice function that Arrow described. They have the same domain but the social choice function applied to B above would produce $(x > y > z)$ so they have different ranges. Let the range of v^{nm} be T_p with $1 \leq p \leq m$ such that p is the cardinality of T_p. Let the tetrad $\langle I_n, S_m, v^{nm}, T_p \rangle$ be the *committee structure*. The committee structure does not affect our voting mechanism, but given that the range of a voting mechanism, unlike a social choice function, does not have a fixed cardinality, it is a useful mechanism for presenting the key characteristics of a voting procedure.

Now, let ψ_W, defined for any $W \subseteq S_m$, be the choice function which maps π_m to non-empty subsets of S_m. ψ_W picks the elements of W which the committee members rank highest. More simply, given a range and a single ballot, ψ finds the most preferred alternative for that voter in the range. Consider the following example:

$$S_m = \{x, y, z\},\ B_i = (x > y > z),\ W = \{y, z\}$$

$$\psi_{\mathbf{W}}(\mathbf{B_i}) = \mathbf{y}$$

The proof to follow will consider a strict voting mechanism. A strict voting mechanism admits only strict orderings of the alternatives. While weak orderings of the alternatives are admissible by both Arrow's impossibility theorem and the Gibbard and Satterthwaite theorem, we conduct our analysis using strict voting mechanisms and then extend our analyses to weak orderings of the alternatives. We now discuss the method of extending those results.

Let ρ_m and ρ_m^n denote the corresponding sets to π_m and π_m^n which are strong orders over the alternatives. If $x, y \in S_m$, $x \neq y$, and $R_i \in \rho_m$, then $x \geq y$ in R_i implies $x > y$ in R_i. Similarly, if $x, y \in S_m$, $x \neq y$, and $B_i \in \rho_m$, then $x \geq y$ in B_i implies $x > y$ in B_i. Using this method (replacing weak relative rankings with strict relative rankings) we can conduct our analyses on strict preferences but extend our results to weak preferences. The method of converting weak preferences to strict preferences may seem reductive and arbitrary but it is key to realize that we are simply choosing a procedure which allows us to take either weak or strict preferences as inputs without changing our results.

Much of the following proof rests on weak and strong alternatives excluding voting mechanisms. We first define weak alternatives excluding here and then define strong alternatives excluding voting mechanisms later.

Definition 13. *Let v^{nm} be a strict voting mechanism. It is weak alternative excluding if there is at least one alternative, $x \in S_m$, such that $v^{nm}(B) \neq x$ for all $B \in \rho_m^n$.*

More simply, v^{nm} is weak alternative excluding if its range is strictly contained in S_m.

We now begin the process of formalizing our definition of a strategy-proof voting mechanism. We say v^{nm} is manipulable at B if, for some ballot set, B, such that an individual, $i \in I_n$, who can

substitute vote B_i' for B_i and change the outcome of the voting procedure. A voting mechanism, v^{nm} is strategy-proof if and only if no $B \in \pi_m^n$ exists so that it is manipulable. In economics terms, if a voting mechanism is strategy-proof, every set of sincere strategies: $R = (R_1, ..., R_n) \in \pi_m^n$ is a Nash equillibria.[1] If the voting mechanism is not strategy-proof, there is a set of strategies $R = (R_1, ..., R_n) \in \pi_m^n$ such that $R_i \in R$ are not Nash equillibria. Given this framework, we can now begin the proof that if a strict voting mechanism, v^{nm}, over at least three alternatives, is strategy-proof, it is dictatorial.

Theorem 2 (Gibbard-Satterthwaite)**.** *Consider a strict commitee with structure $\langle I_n, S_m, v^{nm}, T_p\rangle$ where $n \geq 1$, $m \geq p \geq 3$. The voting mechanism v^{nm} is strategy-proof if and only if it is dictatorial.*

Our proof of the Gibbard-Satterthwaite theorem will begin similarly to our proof of Arrow's impossibility theorem: by establishing the Pareto optimality condition. With this in mind, we reconstruct our definition of Pareto optimality to fit the notation of this section.

Definition 14 (Pareto Optimality)**.** *Consider a strict committee $\langle I_n, S_m, v^{nm}, T = T_p\rangle$. The voting mechanism, v^{nm}, is Pareto optimal if and only if for all $B = (B_1, ..., B_n)$ such that $\psi_T(B_1) = ... = \psi_T(B_n)$, $\psi_T(B_i) = v^{nm}(B)$.*

Lemma 7. *Consider a strict commitee with structure $\langle I_n, S_m, v^{nm}, T = T_p\rangle$ where $n \geq 1$, $m \geq p \geq 3$. If v^{nm} is strategy-proof, it must be Pareto optimal.*

Proof. Assume that v^{nm} is strategy-proof and not Pareto optimal. Then there must exist some optimal choice, $x \in T_p$ and ballot set $C \in \rho_m^n$ such that $\psi_T(C_1) = ... = \psi_T(C_n) \neq v^{nm}(C)$. By definition of ψ_T, $\psi_T(C_1) \in T_p$. Then there must exist ballot set $D \in \rho_m^n$ such that $v_{nm}(D) = \psi_T(C_1)$. If $v^{nm}(C_1, ..., C_n) \neq \psi_T(C_1) = v^{nm}(D_1, ..., D_n)$, there must exist some k such that $v^{nm}(D_1, ..., D_{k-1}, C_k, C_{k+1}, ..., C_n) \neq \psi_T(C_1) = v^{nm}(D_1, ..., D_{k-1}, D_k, C_{k+1}, ...C_n)$. We will refer to voter k as the *pivotal voter*. Let the pivotal voter have preferences $R_k \equiv C_k$. Then $\psi_T(C_1) = \psi_T(C_k)$ is individual k's optimal alternative. It follows that their best strategy would be to vote D_k instead of $C_k \equiv R_k$. So v^{nm} is manipulable and thus not strategy-proof. By contradiction, we conclude that if v^{nm} is strategy-proof, it must be Pareto optimal. □

We will now re-express the strategy of the previous proof in simpler language. We begin by showing that there is, by definition, a set of ballots, D, where $v^{nm}(D)$ will produce the most preferred alternative for voters in ballot set C. Additionally, we show that as we begin replacing ballots C_i with D_i there exists a pivotal voter such that $v^{nm}(D_1, ..., D_{k-1}, C_k, ..., C_n) \neq v^{nm}(D_1, ..., D_{k-1}, D_k, ...C_n)$. But we know that voters in ballot set C prefer $v^{nm}(D)$ to $v^{nm}(C)$ so the consequential voter: C_k has an incentive to cast a ballot, D_k which is different from their sincere preferences, $C_k \equiv R_k$. This is a contradiction because we have constructed a ballot set such that the voting mechanism is not strategy-proof. Next, we define a strong alternative excluding voting procedure.

Definition 15. *A voting mechanism, v^{nm}, is strong alternative excluding if it is both weak alternative excluding and Pareto optimal.*

The next three lemmas focus on a three-element set of alternatives and prove that if we have a strict voting mechanism, $v^{n,3}$, with a range, T_p where $1 \leq p \leq 3$, it must be either dictatorial or strong alternative excluding. Note that if p $<$ 3, it is trivial to show that the voting mechanism is strong alternative excluding. Then, we will focus on proving that if p = 3, the voting mechanism must be dictatorial. The next three lemmas produce an inductive chain over any number of individuals on the committee: n. In particular, the next lemma uses fully dictatorial voting mechanisms. A

[1]Nash equilibrium is a concept in Economics where no agent can improve their outcome without worsening the outcome of another agent

fully dictatorial voting mechanism is one where the range of the voting mechanism is S_m or, by the construction of the voting mechanism, any alternative can be the committee choice provided the right ballot set. In contrast, a partially dictatorial voting mechanism has a range strictly less than the alternative set.

Lemma 8. *Consider a strict committee $\langle I_n, S_m, v^{1,3}, T = T_p\rangle$ where $1 \leq p \leq 3$. If $v^{1,3}$ is strategy-proof it is either fully dictatorial or strong alternative excluding.*

Proof. Towards contradiction, assume there is a strict voting mechanism $v^{1,3}$ which is strategy-proof but neither fully dictatorial nor strong alternative excluding. Then one of the following conditions must be true:

(A) $v^{1,3}$ is Pareto optimal and not weak alternative excluding
(B) $v^{1,3}$ is Pareto optimal and weak alternative excluding
(C) $v^{1,3}$ is not Pareto optimal

The first condition cannot hold because, by assumption, we only have a single voter so if $T_p = S_m$ and our voting mechanism is Pareto optimal, it is necessarily fully dictatorial. The second condition cannot hold because, by definition, if a voting mechanism is weak alternative excluding and Pareto optimal it is strong alternative excluding. The third condition cannot hold because we proved in the previous lemma that every strategy-proof voting mechanism is necessarily Pareto optimal. Then it must be the case that if $v^{1,3}$ is strategy-proof it is either fully dictatorial or strong alternative excluding. □

Lemma 9. *Consider a strict committee $\langle I_n, S_m, v^{n+1,m}, T_p\rangle$ where $n \geq 1$ and $1 \leq p \leq 3$. Let $B = (B_1, ..., B_n)$. The strict voting mechanism, $v^{n+1,3}$, may be written as:*

$$v^{n+1,3}(B, B_{n+1}) = \begin{cases} v_1^{n,3}(B) \text{ if } B_{n+1} = (x\ y\ z) \\ v_2^{n,3}(B) \text{ if } B_{n+1} = (x\ z\ y) \\ ... \\ v_6^{n,3}(B) \text{ if } B_{n+1} = (z\ y\ x) \end{cases}$$

where $v_1^{n,3}, v_2^{n,3}, ..., v_6^{n,3}$ are strict voting mechanisms for n member committees. Then $v^{n+1,m}$ is strategy proof if and only if there is no $(B_n, B_{n+1}) \in \pi_m^n$ such that there exists a $j \in I_n$ who can manipulate v_k^n where $1 \leq k \leq 6$.

This lemma suggests that a strategy-proof voting mechanism can only be constructed from a set of strategy-proof voting mechanisms: $\{v_1^{n,3}, v_2^{n,3}, ..., v_6^{n,3}\}$. If one of the voting mechanisms is not strategy-proof, voter $n+1$ can manipulate the outcome by casting the ballot that corresponds to that sub-voting mechanism. Thus, the broader voting mechanism is not strategy-proof. We now begin the proof of the lemma.

Proof. Towards contradiction, assume that there exists a voting mechanism, $v^{n+1,3}$, which is strategy-proof yet at least one of the 6 voting mechanisms, $v_k^{n,3}$ with $1 \leq k \leq 6$, is not strategy-proof. Without loss of generality assume $v_1^{n+1,3}$ is not strategy proof for individual $j \in I_n$. Then, there must exist some ballot set $B = (B_1, B_2, B_3, ..., B_n) \in \rho_3^n$ such that individual j can manipulate $v_1^{n+1,3}$ at B by casting ballot $(x\ y\ z)$. Let $B' = (B_1, ..., B_i', ..., B_n)$. Then $v^{n+1,3}(B, B_{n+1}) = v_1^{n,3}(B)$ and $v^{n+1,3}(B', B_{n+1}) = v_1^{n,3}(B')$. Therefore voting mechanism $v^{n+1,3}$ is manipulable at (B, B_{n+1}). This is a contradiction because we assumed that $v^{n+1,3}$ was strategy-proof. Therefore $v^{n+1,3}$ is strategy proof if and only if $v_k^{n+1,3}$ are strategy proof for $1 \leq k \leq 6$. □

Note that while all sub-voting mechanisms being strategy-proof is a necessary condition for a strategy-proof voting mechanism, it is not a sufficient condition.

Lemma 10. *Consider a strict committee $\langle I_n, S_m, v^{nm}, T = T_p\rangle$ where $n \geq 1$ and $1 \leq p \leq 3$. Every strategy-proof voting strict procedure, $v^{n,3}$, is either fully dictatorial or strong alternative excluding.*

Proof. Let $\mathcal{V}^n$ and $\mathcal{V}^{n+1}$ be the sets of all strict voting mechanisms, $v^{n,3}$ and $v^{n+1,3}$, for a committees with n and $n+1$ members respectively. Additionally, let $\mathcal{X}^n \subset \mathcal{V}^n$ and $\mathcal{X}^{n+1} \subset \mathcal{V}^{n+1}$ be the sets of strict voting mechanisms which are fully dictatorial or strong alternative excluding. Now let $\mathcal{W}^{n+1} \subset \mathcal{V}^{n+1}$ be the collection of strict voting mechanisms, $v^{n+1,3} \in \mathcal{V}^{n+1}$ which are constructed from voting processes $v^{n,3} \in \mathcal{V}^n$ which are strategy proof. Then, by our previous result,

$$v^{n+1,3}(B, B_{n+1}) = \begin{cases} v_1^{n,3}(B) \text{ if } B_{n+1} = (x\ y\ z) \\ v_2^{n,3}(B) \text{ if } B_{n+1} = (x\ z\ y) \\ ... \\ v_6^{n,3}(B) \text{ if } B_{n+1} = (z\ y\ x) \end{cases}$$

Finally, let $\mathcal{V}^{n*}$ and $\mathcal{V}^{n+1*}$ be the sets of strategy proof voting mechanisms in $\mathcal{V}^n$ and $\mathcal{V}^{n+1}$. If all strategy-proof voting mechanisms are either dictatorial or strong alternative excluding, then $\mathcal{V}^{n*} \subset \mathcal{X}^n$ and, by lemma 6, $\mathcal{V}^{n+1*} \subset \mathcal{W}^{n+1}$. It follows that all $v^{n,3} \in \mathcal{V}^{n+1*}$ can be found by partitioning $\mathcal{W}^{n+1}$ and removing all $v^{n,3} \notin \mathcal{V}^{n+1*}$. We can construct this partition by recognizing that $\mathcal{W}^{n+1}$ can be partitioned into seven subsets. Recall that our alternative set has a cardinality of three. Then $p \leq 3$. So the total number of possibilities can be written as

$$\sum_{p=1}^{3} \binom{3}{p} = \binom{3}{1} + \binom{3}{2} + \binom{3}{3} = 7$$

The final term is the fully dictatorial voting mechanism and the first two are the strong alternative excluding voting mechanisms. Then each subset will be in one of the forms:

$$\mathcal{W}_g^{n+1} = \{v^{n+1,3} \mid v^{n+1,3} \in \mathcal{W}^{n+1} \ \&\ v^{n+1,3}[B, (x\ y\ z)] = h_K^{n,3}\}$$

$$\mathcal{W}_g^{n+1} = \{v^{n+1,3} \mid v^{n+1,3} \in \mathcal{W}^{n+1} \ \&\ v^{n+1,3}[B, (x\ y\ z)] = f_T^{i}\}$$

where h is a strong alternative excluding voting mechanism, f is a dictatorial voting mechanism, $1 \leq g \leq 7$, $K \subseteq S_3$, $T = S_3$, and $i = I_n$. More simply, each voting mechanism must be either strong alternative excluding (the first set) with K denoting the range or fully dictatorial (the second set) where i denotes the dictator in I_n and T asserts that the range of the voting mechanism is all of S_m.

Each of the seven sets can itself be partitioned into seven subsets, $\{W_{1,1}, ..., W_{1,7}, W_{2,1}, ..., W_{7,7}\}$ in the following manner:

$$\mathcal{W}_{a,b}^{n+1} = \{v^{n+1,3} | v^{n+1,3} \in \mathcal{W}_a^{n+1} \ \&\ v^{n+1,3}[B, (x\ y\ z)] = h_K^{n,3}\}$$

Here $W_{a,b}^{n+1}$ denotes the set of strict voting mechanisms such that $v^{n+1,3} \in \mathcal{W}_a^{n+1}$ and b denotes which of the 7 sets (the 6 strong alternate excluding and the fully dictatorial) set corresponds to $v^{n+1,3}[B, (x\, y\, z)]$. Many of these can be shown to be disjoint from $\mathcal{V}^{n+1*}$ by considering cases where the individual's most preferred alternative is not in the range of the voting mechanism. If this is the

case the individual can benefit from casting a ballot with an alternative different from their most preferred ranked highest. After a similar process for all the subsets, Gibbard and Satterthwaite find that seventeen subsets are not disjoint with $\mathcal{V}^{n+1*}$ and all are alternative excluding or fully dictatorial. Then $\mathcal{V}^{n+1*} = (\mathcal{V}^{n+1*} \cap \mathcal{W}^{n+1}) \subset \mathcal{X}^{n+1}$. □

Lemma 8 produces an inductive chain that shows that if a strict voting mechanism is strategy-proof it is either fully dictatorial or strong alternative excluding for any size committee. A similar inductive chain can be constructed on the number of alternatives. While I could not find Gibbard and Satterthwaite's proof of that inductive chain, I propose the following proof.

Lemma 11. *Let $v^{n,m}$ be either fully dictatorial or strong alternative excluding. Then $v^{n,m+1}$ is fully dictatorial or strong alternative excluding as well.*

Proof. First consider the case where $v^{n,m}$ is strong alternative excluding. This, by definition, means that $T_p \subset S_m$. Therefore there exists some $x \in S_m$ such that $x \notin T_p$. Because $S_m \subset S_{m+1}$, $x \in S_m$. It follows that $x \in S_{m+1}$. Then there exists $x \in S_{m+1}$ such that $x \notin T_p$. Therefore $v^{n,m+1}$ is strong alternative excluding as well. Thus we have shown if $v^{n,m}$ is strong alternative excluding, then $v^{n,m+1}$ is strong alternative excluding.

Next consider the case where $v^{n,m}$ is fully dictatorial. First, consider the case where the new alternative is not in the range of the voting mechanism. Then the voting mechanism is once again strong alternative excluding.

Finally consider the case where $v^{n,m}$ is fully dictatorial and the range of $v^{n,m+1}$ is S_{m+1}. Without loss of generality assume that voter k was the dictator in $v^{n,m}$. Then voter k controlled the societal relative ranking of all m alternatives. Towards contradiction, assume that $v^{n,m+1}$ is not dictatorial. Then there exists no voter who controls the relative ranking of alternatives. This means that the introduction of a new alternative changed the method of determining the rankings of the other m alternatives. But this violates independence of irrelevant alternatives. Then voter k must remain the dictator.[2] □

Theorem 3. *Consider a strict committee $\langle I_n, S_m, v^{nm}, T = T_p \rangle$ where $n \geq 1$ and $p \geq 1$. If v^{nm} is strategy-proof, then it is either fully dictatorial or strong alternative excluding.*

Proof. This is a generalization of the previous few results. Lemma 6 showed that for a committee with one person and three alternatives, a voting mechanism is either fully dictatorial or strong alternative excluding. Lemma 8 constructed an inductive chain on the number of voters. Additionally, Gibbard and Satterthwaite asserted that a similar inductive chain can be generated on the number of alternatives. Then, as this lemma states, we can extend the results of the previous lemmas to any finite number of alternatives or members of the voting committee. □

Before the next lemma, we define a new function. Let θ_W be a mapping from π_m to π_q such that if $x, y \in W$, $C_i \in \pi_q, D_i \in \pi_m$, and $C_i = \theta_W(D_i)$, then $x > y$ in C_i if and only if $x > y$ in D_i. The function θ constructs a new weak ordering, C_i, from D_i by deleting the elements of S_m which are not in the range, W.

[2]This result should be taken as given by the original authors. I offer my proof just so the reader can understand the intuition about why such an inductive chain can be constructed. I feel confident about my proof for the first two cases. I assume, though I do not know, that the author's proof of the third case will reduce to the argument I have presented but will make the argument more rigorous. Note that the intuition I used in my proof draws from our proof of Arrow's impossibility theorem where we found that if one individual controls the relative ranking of two alternatives, they control the relative rankings of all alternatives.

Lemma 12. *Consider a strict committee $\langle I_n, S_m, v^{nm}, T = T_p\rangle$ where $n \geq 1$, $m \geq 3$, $p \geq 1$ and $m \geq p$. If v^{nm} is a strategy-proof voting mechanism and two ballot sets, $C, D \in \rho_m^n$ have the property that for all $i \in I_n$, $\theta_T(C_i) = \theta_T(D_i)$, then $v^{nm}(C) = v^{nm}(D)$.*

Proof. Note that the final line in the lemma asserts that the ordinal rankings of the alternatives must be the same between the two sets of ballots. Additionally, note that if $T = S_m$ this result must trivially hold because our restriction guarantees that C is identical to D. The more interesting case occurs when $T \subset S_m$. Then assume that v^{nm} is strategy-proof but not alternative excluding. Note that this contradicts Proposition 1. There must exist two sets of ballots, $C, D \in \rho_m^n$ such that $v^{nm}(C) \neq v^{nm}(D)$ and for all $i \in I_n$ $\theta_T(C_i) = \theta_T(D_i)$. More simply, there must be two sets of ballots so that their ordinal rankings are the same but the alternative chosen by the strict strategy-proof voting mechanism differs. Now consider, just as we did previously, the sequence of ballots where we begin replacing ballots in set C with ballots from set D. There must exist a pivotal voter, k, and two alternatives, $x \neq y \in S_m$ such that $v^{nm}(D_1, ..., D_{k-1}, C_k, ..., C_n) = x\grave{n}ey = v^{nm}(D_1, ..., D_{k-1}, D_k, C_{k+1}, ..., C_n)$. Because we are evaluating strict committees, we have removed indifference. So either x is preferred to y in both ballots C and D or y is preferred to x in both ballots C and D. In both cases, there is an opportunity for individual i to manipulate the outcome, which contradicts this voting mechanism being strategy-proof. □

We are now ready to complete the proof of theorem 2 which asserts that a voting mechanism v^{nm} with at least three alternatives is strategy-proof if and only if it is dictatorial. Recall that being dictatorial, by inspection, implies that the voting mechanism is strategy-proof. The dictator is incentivized to vote according to their sincere strategy because they determine the outcome and would like to have their most preferred outcome selected. All other voters have no motivation to adopt a sophisticated strategy because, by doing so, they can never produce a better outcome by the standards of their sincere preferences. Then to complete this proof we must show the converse. We have already shown that if a strict voting mechanism, v^{nm} is strategy-proof it is either fully dictatorial or strong alternative excluding. So we need to show that if v^{nm} is both strategy-proof and strong alternative excluding, it must be partially dictatorial.

Theorem (Gibbard-Satterthwaite)**.** *Consider a strict committee with structure $\langle I_n, S_m, v^{nm}, T_p\rangle$ where $n \geq 2$, $m \geq p \geq 3$. The voting mechanism v^{nm} is strategy-proof if and only if it is dictatorial.*

Proof. Assume that a voting mechanism, v^{nm} is strategy proof and is strong alternative excluding with range $T = T_p$, and $m > p \geq 3$. Now replace all ballots $B_i \in \rho_m^n$ with a strong ordering, $B_i^* \in \rho_p^n$ where $B_i^* = \theta_T(B_i)$. Consider any $C, D \in \rho_m^n$ such that $C \neq D$ and $[\theta_T(C_1), ..., \theta_T(C_n)] \neq [\theta_T(D_1), ..., \theta_T(D_n)]$. By lemma 12 $v^{nm}(C) = v^{nm}(D)$. There exists a strict voting mechanism, v^{np} such that $v^{np}[\theta_T(C_1), ..., \theta_T(C_n)] = v^{nm}(B_1, ..., B_n)$. Because v^{nm} is strategy proof, v^{np} must be as well. Therefore v^{np} must be either fully dictatorial or strong alternative excluding. But it cannot be strong alternative excluding because its range includes all elements of T_p. Therefore it must be partially dictatorial. □

Therefore a voting procedure, v^{nm} in strict committee with structure $\langle I_n, S_m, v^{nm}, T_p\rangle$ where $n \geq 1$, $m \geq p \geq 3$ is strategy proof if and only if it is dictatorial. This completes the proof of the Gibbard-Satterthwaite theorem. ■

The Gibbard-Satterthwaite theorem concludes that all non-dictatorial voting mechanisms provide incentives for at least one ballot set such that voters reveal preferences other than their own, and the resulting social choice may then be distorted away from the Pareto optimum relative to their true tastes. Notice that although we never invoked the result of Arrow's Impossibility Theorem, we arrived at the same result for strategy-proof voting mechanisms. The next section makes this

connection clearer. It shows that there is a one-to-one correspondence between strategy-proof voting mechanisms and social choice functions.

3.2. **From Strategy Proof Voting Mechanisms to Arrow's Impossibility Theorem.** In this section, we seek to present the correspondence that Gibbard and Satterthwaite developed between Arrow's Impossibility Theorem and the Gibbard-Satterthwaite Theorem. This section begins by restating the conditions for Arrow's Impossibility Theorem. Then, we develop the connection between a strict social choice function and a strict voting mechanism. [3] Next, we prove that any strict social choice function satisfying Arrow's impossibility theorem can be turned into a strict voting mechanism. We then show that the converse is true. Finally, we prove that every social choice function that constructs a strict voting mechanism and vice versa is unique. By showing the bi-conditional implication we establish a bijection between the set of strict social choice functions and the set of strategy-proof voting mechanisms.

Recall that Arrow's Impossibility Theorem defined a social choice function for a committee of n members and m alternatives as a mapping, which we will now denote u^{nm}, with domain π_m^n and range π_m. Then $u^{nm}(B) = A$ where $B = (B_1, B_2, ..., B_n) \in \pi_m^n$ and $A \in \pi_m$. A is a weak ordering that is called the *social choice*. As we noted in the previous proof, the difference between the social choice function and the strict voting mechanism is that the range of the social choice function is a weak ordering of the alternatives instead of a single alternative. Given a particular ballot set and range, the *committee choice* is the highest-ranked alternative. Let a committee using the social choice function u^{nm} be described using the triplet $\langle I_n, S_m, u^{nm}\rangle$.

Recall from the first section that Arrow requires the social choice function to have the qualities of non-dictatorship, independence of irrelevant alternatives, and monotonicity. While Arrow also required rationality we can now assume that social choice functions and strict voting mechanisms are rational. As we previously discovered, these conditions implicitly require our social choice function to abide by Pareto optimality.

We begin by describing how a strategy-proof voting mechanism can be constructed from any social choice function satisfying Arrow's conditions. Let u^{nm} be a social choice function with the property that for all $B \in \pi_m^n$, $\psi_s(u^{nm}(B))$ is always a single element of S_m. Now let v^{nm} be defined so that for any $B \in \pi_m^n$, $v^{nm}(B) = \psi_S[u^{nm}(B)]$. We can call any v^{nm} a voting mechanism derived from u^{nm} so that we can map the u^{nm} to v^{nm}. Because u^{nm} satisfies the conditions for Arrow's impossibility theorem, v^{nm} does as well. More intuitively, we can find a voting mechanism by constructing a social choice function, deriving the social rankings, and selecting the highest-ranked alternative. To visualize this, consider the example below where $n = 5$, $m = 3$, and $S_m = \{x, y, z\}$:

$$u^{5,3}(B) = u^{5,3}\begin{pmatrix} B_1: & x > y > z \\ B_2: & x > y > z \\ B_3: & x > y > z \\ B_4: & x > y > z \\ B_5: & x > y > z \end{pmatrix} = (x > y > z) \in \pi_m$$

$$\mathbf{v^{nm}} = \psi_{\mathbf{S}}(\mathbf{u^{nm}}) = \mathbf{x} \in \mathbf{S_m}$$

See that the social choice function constructs the societal ranking of alternatives and then the voting mechanism selects the highest-ranked alternative from the social choice. We must still show that the voting mechanisms constructed from the social choice functions are strategy-proof.

[3]During the proof of Arrow's Impossibility Theorem in section 1 we referred to the social choice function as a Constitution

Lemma 13. *Consider a strict committee $\langle I_n, S_m, u^{nm}\rangle$ where $n \geq 2$, $m \geq 3$. If the strict social choice function satisfies Arrow's conditions, then the strict voting mechanism v^{nm} derived from u^{nm} is strategy-proof and range $T \equiv S_m$.*

Proof. Recall that if u^{nm} satisfies Arrow's conditions, it is Pareto optimal. Note that if u^{nm} is Pareto optimal, v^{nm} must have range identical to S_m because u^{nm} has domain ρ_m^n and $v^{nm} = \psi_S[u^{nm}(B)]$ for all $B \in \rho_m^n$. Then we have shown that v^{nm} has range $T \equiv S_m$. Next, we must show that v^{nm} is strategy proof.

Towards contradiction, assume u^{nm} meets Arrow's conditions but v^{nm} derived from u^{nm} is not strategy proof. By definition of strategy-proofness, there must exist a ballot set, $B = (B_1, B_2, ...B_k, ..., B_n) \in \rho_m^n$ such that $v^{nm}(B_1, B_2, ...B_k, ..., B_n) = y \neq x = v^{nm}(B_1, B_2, ...B'_k, ..., B_n)$ and $x > y \in R_k$. Assume $u^{nm}(B_1, B_2, ...B_k, ..., B_n) = A \in \rho_m$ and $u^{nm}(B_1, B_2, ...B'_k, ..., B_n) = A' \in \rho_m$. Note that, by definition of ψ, $\psi_S(A') = x$ and $\psi_S(A) = y$. There are two cases in B'_k: either x is preferred to y or y is preferred to x. We first consider the case where y is preferred to x.

Let $U = S_m - (x)$ or all other alternatives in the range except x. Now let ballot $B_k^* = [x\, \theta_U(B'_k)]$ so that B_k^* is the ballot B_k where x is made the most highly ranked alternative and the rankings of all other alternatives are unchanged. Let $u^{nm}(B_1, B_2, ..., B_k^*, ..., B_n) = A^*$. Recall that $\psi_S(A') = x$ so $\psi_S(A^*) = x$. Then x must be the most preferred alternative in A^*. Let $X = \{x, y\}$ and note that, by construction of B_k^*, $\theta_X(B_i^*) = \theta_X(B_i)$ or x is preferred to y in both B_i and B_i^*. By independence of irrelevant alternatives, $\psi_X(A^*) = \psi_X(A)$ but this is a contradiction because $\psi_X(A^*) = \psi_S(A^*) = x \neq y = \psi_X(A) = \psi_S(A)$. Thus if y is preferred to x, the voting mechanism must be strategy-proof.

Now consider the case where x is preferred to y under B'_k. Again observe that $\theta_X(B_k) = \theta_X(B'_k)$ where $X = \{x, y\}$. By independence of irrelevant alternatives, $\psi_X(A') = \psi_S(A)$. But this once again contradicts $\psi_X(A') = \psi_S(A') = x \neq y = \psi_X(A) = \psi_S(A)$. Therefore the voting mechanism must be strategy-proof. □

In this proof, we moved elements in individuals' rankings while preserving their relative rankings and showed that, using that process, we could change the outcome of the voting procedure. That provided a contradiction because we assumed the voting procedure was strategy-proof. That contradiction allowed us to show that our social choice function must obey Arrow's conditions. We have now completed showing that a strict mechanism can be constructed from a social choice function and that the voting mechanism is strategy-proof. Next, we must show that we can construct a social choice function from a voting mechanism.

Lemma 14. *Consider a strict committee $\langle I_n, S_m, v^{nm}, T_p\rangle$ where $n \geq 2$, $m \geq 3$, and $T_p \equiv S_m$. If v^{nm} is strategy-proof there exists a unique, strict social choice function, u^{nm}, which underlies v^{nm} and satisfies Arrow's conditions.*

Proof. First, we need to show that for all strict strategy-proof voting mechanisms there is a strict social choice function. Let $Q \in \rho_m$ be an arbitrary strong ordering over the alternatives. Define λ_{xy} where $x, y \in S_m$ and $x \neq y$ to be a function with both the domain and range ρ_m. Additionally let λ_{xy} have the following properties such that if $B_k^* = \lambda_{xy}(B_k)$ then

(A) x is preferred to y under B_k^* if x is preferred to y under B_k
(B) y is preferred to x under B_k^* if y is preferred to x under B_k
(C) x and y are preferred to w under B_k^* for all $w \in S_m - (x, y)$
(D) w is preferred to z under B_k^* if w is ranked higher than z in Q for all $w, z \in S_m - (x, y)$.

Now for each ballot set $(B_1, B_2, ..., B_n)$ construct a binary relation, P such that, for all $x, y \in S_m$, x is preferred to y in P if and only if $x = v^{nm}[\lambda_{xy}(B_1), ..., \lambda_{xy}(B_n)]$. More simply, x is preferred

to y in P if and only if the majority of voters prefer x to y. Because P is defined for all $B \in \rho_m^n$, a function can be defined, μ that associates each P with each $B \in \rho_m$. We showed earlier that if a strict voting mechanism is strategy-proof, the binary relation P associated with each $B \in \rho_m^n$ must be a strong order: $P \in \rho_m$. Then μ is a strict social choice function. □

To visualize this consider the following ballot set for $m = n = 3$, $S_m = \{x, y, z\}$:

$$B_1 = (x > y > z) \qquad \lambda_{x,y}(B_1) = (x > y > z)$$
$$B_2 = (y > x > z) \qquad \lambda_{x,y}(B_2) = (y > x > z)$$
$$B_3 = (z > x > y) \qquad \lambda_{x,y}(B_3) = (x > y > z)$$

See that the function λ moves alternatives x and y to the first and second positions in the rankings and chooses their relative ranking based on the individual ballot. Because the voting mechanism only looks at the most preferred alternative, to construct relative rankings between any two alternatives we must have one of the two be the most ranked alternative for all ballots. Then, if we continue this process for all pairwise sets alternatives we can construct a complete ranking of all alternatives. That process is shown below:

$$v^{nm}(\lambda_{x,y}(B_1), \lambda_{x,y}(B_2), \lambda_{x,y}(B_1)) = x$$
$$\mathbf{x > y}$$
$$v^{nm}(\lambda_{y,z}(B_1), \lambda_{y,z}(B_2), \lambda_{y,z}(B_1)) = y$$
$$\mathbf{y > z}$$
$$v^{nm}(\lambda_{x,z}(B_1), \lambda_{x,z}(B_2), \lambda_{x,z}(B_1)) = x$$
$$\mathbf{x > z}$$

$$\mathbf{P} = (\mathbf{x > y > z})$$

When we aggregate each of these relative rankings we construct P, a societal ranking of the alternatives. Then μ maps ballot sets to societal rankings of the alternatives. It follows, by definition of a social choice function, that μ is a social choice function.

Note that although the framework we have employed here considered only strict voting mechanisms with ranges T_p equivalent to the alternative set S_m, this condition is not limiting as we showed previously, any strict strategy-proof voting mechanism v^{nm} with a range $T_p \subset S_m$, $p \geq 3$, can be written as a strategy-proof voting mechanism, v^{np} defined over the reduced alternative set $S_p = T_p$. So the result applies to v^{np} which has an underlying strict social choice function u^{np} which satisfies Pareto optimality and independence of irrelevant alternatives. Additionally, note that our result does not establish that u^{nm}, which underlies v^{nm}, is unique. We will now prove that fact.

Lemma 15. *Consider a strict committee $\langle I_n, S_m, v^{nm}, T_p \rangle$ where $n \geq 1$, $m \geq 3$, and $T_p \equiv S_m$. If v^{nm} is strategy-proof, there exists a unique, strict social choice function u^{nm} which underlies v^{nm} and satisfies Arrow's conditions.*

Proof. Suppose two strict social choice functions, μ and μ' both underlie v^{nm}, both satisfy Pareto optimality and independence of irrelevant alternatives, and for some ballot set, C, $\mu(C) \neq \mu'(C)$. Note that because both μ and μ' underlie v^{nm}, for all $B \in \rho_m^n$ $v^{nm}(B) = \psi_S[\mu(B)] = \psi_S[\mu'(B)]$. Then there exists $x, y \in S_m$ such that x is preferred to y under A and y is preferred to x under A' where $\mu(C) = A$ and $\mu'(C) = A'$. Let $C_i^* = \lambda_{xy}(C_i)$ for all $i \in I_n$. Additionally, let $A^* = \mu(C^*)$, and $A^{'*} = \mu'(C^*)$. By independence of irrelevant alternatives, x is preferred to y under A^* and y is preferred to x under $A^{'*}$. By Pareto optimality x and y are preferred to z under both A^* and $A^{'*}$

for all $z \in S_m - \{x, y\}$. Therefore $\psi_S(A^*) = \psi_S[\mu(C^*)] = x \neq y = \psi_S[\mu'(C^*)] = \psi_S(A'^*)$. This is a contradiction because we assumed that $\psi_S[\mu(B)] = \psi_S[\mu'(B)] = v^{nm}(B)$ for all $B \in \rho_m^n$. Then it must be the case that $\mu = \mu'$ or that there is only one social choice function that underlies each strict strategy-proof voting mechanism.

Assume that u^{nm} satisfies Pareto optimality and independence of irrelevant alternatives, but not monotonicity. Then there must exist $x, y \in T_p$, $B = (B_1, ..., B_k, ..., B_n) \in \rho_m^n$, and $B'_k \in \rho_m$ such that

(A) y is preferred to x under B_k
(B) x is preferred to y under B'_k
(C) y is preferred to x under A'
(D) x is preferred to y under A

where $A = u^{nm}(B_1, ..., B_k, ..., B_n)$ and $A' = u^{nm}(B_1, ..., B'_k, ..., B_n)$. For all $j \in I_n$ assume $C_j = \lambda_{xy}(B_j)$ and $C'_k = \lambda_{xy}(B'_k)$. Because u^{nm} satisfies Pareto optimality and independence of irrelevant alternatives, $\psi_S[u^{nm}(C_1, ...C_k, ..., C_n)] = x \neq y = \psi_S[u^{nm}(C_1, ...C'_k, ..., C_n)]$. Because u^{nm} underlies v^{nm}, $\psi_S[u^{nm}(B)] = v^{nm}(B)$. Then $v^{nm}(C_1, ...C_k, ..., C_n) = x \neq y = v^{nm}(C_1, ...C'_k, ..., C_n)$. Because y is preferred to x and C_k voter k can manipulate v^{nm}. So if u^{nm} does not follow monotonicity, v^{nm} is not strategy-proof. Thus u^{nm} must abide by monotonicity. □

This proof once again proceeded by moving elements in individual ballots and showing that the outcome of the voting procedure can be manipulated using that process—contradicting strategy proofness. That contradiction revealed that the mapping between voting mechanisms and social choice functions must be unique or one-to-one.

Theorem (Gibbard-Sattethwaite Correspondence)**.** *Let $n \geq 1$ and $m \geq 3$. A one-to-one correspondence, γ exists between every strict strategy-proof voting mechanism, v^{nm} with range $T_p \equiv S_m$ and every strict social choice function, u^{nm} satisfying Arrow's conditions. If $u^{nm} = \gamma(v^{nm})$, u^{nm} underlies v^{nm} and v^{nm} is derived from u^{nm}.*

We have shown that the mapping between the set of social choice functions and voting mechanisms is both injective and surjective. Additionally, we have shown that these mappings preserve the characteristics of social choice functions and voting mechanisms that both Arrow, and Gibbard and Satterthwaite, require. Therefore we have proven the Gibbard-Satterthwaite correspondence theorem. ■

4. Extensions of Voting Mechanisms

The previous two sections have been mathematically joyous but bleak in their optimism about inducing individuals to reveal their true preferences. In the final section of this paper, we aim to less rigorously consider some alternative voting mechanisms which provide some short-lived optimism about mechanisms where individuals' dominant strategy is to reveal their true preferences.

One area where the implementation of strategy-proof voting mechanisms would be useful is public goods. Public goods like national defense, streetlights, and many forms of infrastructure, are both non-excludable and non-rivalrous. Individuals have incentives to under-represent their value for public goods because they do not want to carry the burden of paying for the public good and know that once it has been purchased they can benefit from it even if they didn't pay for it.

Additionally, the allocation of contracts to build public goods is often done through auction mechanisms. For example, many bureaucratic agencies conduct sealed lowest unique bid auctions. In these auctions, all agents place a bid on a project (the amount they expect the government to

pay them for completing the job) and the lowest bid wins. But agents in these auctions, seeking to generate a surplus for themselves, often choose cheaper supplies or less skilled labor. The consequences of a poorly constructed bridge or building are significant. Then, having a strategy-proof voting mechanism to uncover people's preferences and coordinate auctions is important. The Vickrey-Clarke-Groves mechanism provides a method for both. [8] The general structure for all VCG mechanisms is:

(A) All agents are asked to report value functions: $v_i(x)$ for all x
(B) The optimal choice is chosen as $x^* = x^{opt}(v)$
(C) Each agent is paid (r_i) according to the following policy:

$$r_i = p_i + h_i(v_{-i})$$
$$p_i = \sum_{j \neq i} (v_j(x^*))$$
$$v_{-i} = (v_1, v_2, ..., v_{i-1}, v_{i+1}, ...v_n)$$

Each agent's receipt (r_i) is the sum of a direct payment (p_i) which is equivalent to the total value of the chosen alternative to all other agents and the value of some arbitrary function $(h(v_{-i}))$ where v_{-i} is the value to all other agents.[4] For example $h(v_{-i}) = 0$ is one option for the function h. But this isn't often used because then $r_i > 0$ in many cases.

$$r_i = p_i + h_i(v_{-i}),\ p_i \geq 0, h_i(v_{-i}) = 0 \implies r_i \geq 0$$

A function of that form is realistic if we have $p_i = \sum_{j \neq i}(v_j(x^*)) < 0$ for all i. Consider the case where a public nuisance (a toxic factory, a waste disposal plant, a loud airport, etc.) is being built and various parties are bidding to have the nuisance built away from them.[5] Then, presumably, $p_i < 0$ for all i. In this case, $h_i(v_{-i}) = 0$ will work as a function for h. More commonly h takes the form of a function like $h_i(v_{-i}) = -\max_{x \in X} \sum_{j \neq i}(v_j(x))$. This is called the Clarke pivot rule and it creates a payout so that every agent pays a value equivalent to:

(social welfare of others if agent (i) was absent) - (social welfare of others when agent (i) is present)

In economics terms, every agent pays their externality. One application of Vickrey-Clarke-Groves mechanisms is Vickrey auctions. In a Vickrey auction, there is a single item which people are bidding on. Additionally, there are $n+1$ outcomes. Each of the first n outcomes denotes that the good goes to one of the n bidders. The last outcome is that the item is not sold to anyone. In the case of the Vickrey auction:

$$p_i = \begin{cases} 0 \text{ if } \mathrm{v}_i > v_j \text{ for all } \ j \neq i \\ \mathrm{v}_i \text{ otherwise} \end{cases} \qquad h_i(v_{-i}) = \begin{cases} -(v_k) \text{ if } \mathrm{v}_i > v_k > v_j \text{ for all } \ j \neq i, k \\ -(\mathrm{v}_j) \text{ otherwise} \end{cases}$$

This reduces to the following conditions:

- All bidders except the highest bidder pay nothing

[4]This is intentionally vague. The broad VCG mechanism is widely applicable so its definition is vague.
[5]This construction may seem strange but it is the same as trying to determine who would be hurt least by this nuisance (who values the right to not have it built next to them the least)

- The highest bidder pays the second-highest bid.

The Vickrey auction is strategy-proof. Bidders never have an incentive to overbid: if they bid higher than their value for the good, they are better off without the good. Because it is a sealed bid auction, raising their bid also does not incentivize others to change their bid. In non-Vickrey auctions bidders typically under-represent their value for the good to extract surplus. If a particular voter, i, values a good at v_i, they hope to buy the item for some price, s_i, where $s_i < v_i$ so that they get both the good and the 'surplus,' $v_i - s_i$. Then bidding v_i in a traditional auction ensures that their payout is either 0 (if they lose the auction) or 0 if they win because they pay exactly what they value the good at. They are engaging in a zero-sum game against themselves. By reducing their bid, the payout of losing remains 0 while the payout (k_i) of winning grows. We can model the expected payout to the bidder as follows:

$$\begin{aligned} k_i &= p(v_i - s_i) + (1-p)(0) \\ &= p(v_i - s_i) \end{aligned}$$

Where p is the probability of winning so p is a function of s_i and $\frac{dp}{ds_i} < 0$. Then there exists a range where $\frac{\delta k_i}{\delta s_i} < 0$ and the bidder has an incentive to under-represent their value for the good. Under the Vickrey auction, the highest bidder does not pay their winning bid so they retain surplus: $v_i - s_j$ where v_i is their value for the good and s_j is the second highest bid. Now the dominant strategy becomes to reveal true preferences or to reveal one's true value of the good being auctioned. There are other types of manipulations that Vickrey auctions give rise to—we discuss those later—but, even in those manipulations, an individual is incentivized to reveal their true value for the auctioned good.

We now turn our attention to Vickrey-Clarke-Groves mechanisms for public goods. Vickrey-Clarke-Groves mechanisms can be applied to public goods by using the Clarke pivot just as it is provided above. Assume the cost of the project is C and each voter, $j \in I_n$, has a value for the project v_j. Under the VCG mechanism, the only bidder who has a non-zero tax for the project is the pivotal bidder (v_i) such that

$$\sum_{j \neq i} (v_j(x^*)) < C$$

$$\sum_{j} (v_j(x^*)) \geq C$$

Here the pivotal voter is the voter without whom the total value of the good to society falls below the cost, but with them, the total value is greater than or equal to the cost. While this does correct the incentives so that individuals reflect their true preferences (because each individual faces such a small probability of having to pay), it has a fatal flaw: it is not tax-neutral. Because only one individual pays their value for the public good, the government does not collect enough tax revenue to sponsor the project and thus must seek other funding sources. This is a significant flaw because understanding individuals' sincere preferences is a futile exercise if they are not induced to act according to those preferences.

Note additionally that the Vickrey auction loses efficacy when a group of individuals collude (such that they collectively can place a very large bid knowing they will not have to pay it). In this sense, Vickrey-Clarke-Groves mechanisms are better in theory than in practical applications. But they reveal something valuable—when we remove the restrictions of Arrow's impossibility theorem and the Gibbard-Sattherthwaite theorem (such as the case where individuals do not need to cast weak order preferences or a single ballot), we can produce voting mechanisms which induce people to reveal

their true preferences. In particular, the Vickrey-Clarke-Groves mechanism has two key qualities: stating one's true preferences is a dominant strategy for each individual and a Pareto optimum is selected. Green and Laffort (1977) proved that the Vickrey-Clarke-Groves mechanism is the only class of mechanisms with these properties. [2] An appropriate examination of their proof would require another paper but it is sufficient for now to point to the specific form of the Vickrey-Clarke-Groves mechanism to explain the intuition. The Vickrey-Clarke-Groves assumes quasi-linear utility functions and therefore their utility is additively separable. More simply it assumes that individuals' utility from a payment and from winning the auction are separable. Then, if an individual wins the auction and gathers a direct payment of $20 (this can also come in the form of a surplus from their bid) their total utility is the utility from the $20 and the utility from the good. Additionally, the total utility they get from that $20 surplus is the same amount of utility they would get if their direct payment in the Vickrey auction was $20 and they did not win the auction. If utility is not additively separable, constructing a mechanism where the dominant strategy is to reveal one's true preferences becomes, as Green and Laffont show, impossible. As a final thought, cardinal voting schemes (such as ones where individuals have a fixed amount of votes to allocate between alternates) are subject to the same result as ordinal voting schemes in the Gibbard-Satterthwaite theorem. We will leave the proof of this final result as an exercise to the reader—after all, leaving an exercise to the reader is the dominant strategy for every mathematician.

References

[1] John Geanakoplos. *Arrow's Paradox*. en. URL: https://math.uchicago.edu/~may/VIGRE/VIGRE2009/REUPapers/Nadathur.pdf.

[2] Jerry Green and Jean-Jacques Laffont. "Characterization of Satisfactory Mechanisms for the Revelation of Preferences for Public Goods". In: *Econometrica* 45.2 (1977). Publisher: [Wiley, Econometric Society], pp. 427–438. ISSN: 0012-9682. DOI: 10.2307/1911219. URL: https://www.jstor.org/stable/1911219 (visited on 11/19/2023).

[3] *How People Voted in Ancient Elections*. en. Nov. 2022. URL: https://www.history.com/news/ancient-elections-voting (visited on 11/19/2023).

[4] George C. Edwards III. *Why the Electoral College Is Bad for America*. en. Google-Books-ID: KU7XEAAAQBAJ. Cambridge University Press, Nov. 2023. ISBN: 978-1-00-942626-8.

[5] Gallup Inc. *Party Affiliation*. en. Section: In Depth: Topics A to Z. Sept. 2007. URL: https://news.gallup.com/poll/15370/Party-Affiliation.aspx (visited on 11/19/2023).

[6] Prerna Nadathur. *Arrow's Paradox*. en. URL: https://math.uchicago.edu/~may/VIGRE/VIGRE2009/REUPapers/Nadathur.pdf.

[7] Mark Allen Satterthwaite. "Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions". In: *Journal of Economic Theory* 10.2 (Apr. 1975), pp. 187–217. ISSN: 0022-0531. DOI: 10.1016/0022-0531(75)90050-2. URL: https://www.sciencedirect.com/science/article/pii/0022053175900502 (visited on 11/19/2023).

[8] *Vickrey–Clarke–Groves mechanism*. en. Page Version ID: 1177931653. Sept. 2023. URL: https://en.wikipedia.org/w/index.php?title=Vickrey%E2%80%93Clarke%E2%80%93Groves_mechanism&oldid=1177931653 (visited on 11/19/2023).