A Simple Proof of the Gibbard-Satterthwaite Theorem

Andrew C. Eggers*

November 13, 2015

ABSTRACT

The Gibbard-Satterthwaite Theorem is a fundamental result in social choice theory: it proves that minimally inclusive and strategy-proof social choice functions do not exist. I offer a simple proof and illustrate the intuition with a new approach to visualizing preferences of three individuals over three alternatives.

Suppose three individuals must collectively choose from among three alternatives. Each individual is asked to submit a ballot stating her order of preference among the three alternatives; an impartial rule will be applied to these ballots to produce a collective choice. Is there a rule that will always induce the voters to submit sincere preference orderings?

Gibbard (1973) and Satterthwaite (1975) proved that the answer is no: apart from the unattractive approach of simply choosing one person and implementing her preferences, all social choice functions reward strategic voting in some circumstances. It had long been known that strategic misrepresentation afflicts specific social choice functions such as plurality rule or the Borda count, and it had been conjectured (Vickrey, 1960; Dummett and Farquharson, 1961) that all minimally inclusive social choice functions suffer from the same problem. Gibbard (1973) and Satterthwaite (1975) proved that this conjecture was correct.

The Gibbard-Satterthwaite Theorem is justly regarded as a core result in social choice theory and it (along with the closely related Arrow Theorem) remains widely taught not just in technical courses on the theory of voting but also as the motivation for studying electoral systems and legislative institutions. And yet it is difficult to find simple and intuitive proofs (or even non-technical explanations) of this fundamental result. Some authors state the result without providing a proof (e.g. Shepsle, 2010); elsewhere it is treated as an extension of Arrow's Theorem (Mueller, 2003). Short and elegant proofs exist (e.g. Barberá, 1983; Reny, 2001), but they tend to be dense and unintuitive for non-specialists.

*First version: October 20, 2015. This version: November 13, 2015. Author contact information: Nuffield College and Department of Politics and International Relations, University of Oxford, UK. email: andrew.eggers@nuffield.ox.ac.uk

This paper provides a new and more accessible proof of the Gibbard-Satterthwaite Theorem. After dealing with some preliminaries, I establish a surprising lemma that can be used directly to prove the Gibbard-Satterthwaite Theorem for the special case of three alternatives and three individuals; for the general case, I use this lemma in conjunction with the method introduced in Geanakoplos (2005)'s proof (and adapted for social choice functions by Reny, 2001). Finally, to solidify the intuition of the proof I provide a graphical explication that introduces a new method for depicting preference profiles in two dimensions.

1. PRELIMINARIES

A *profile* indicates the strict and transitive preference orderings for each member of a society over a set of alternatives. Given three individuals and three alternatives a, b, and c, an example of a profile is

$$\begin{array}{ccc} a & a & c \\ b & c & a \\ c & b & b \end{array},$$

where the first individual prefers a to b and b to c, the second prefers a to c and c to b, etc.

A *social choice function* yields a particular alternative (the social choice) for every possible profile. A *dictatorial* social choice function is one that simply chooses the alternative that is ranked first by a particular individual. A social choice function is *strategy-proof* if there are no circumstances in which an individual would benefit by reporting a preference ordering other than her true preference ordering. A social choice function respects *citizen sovereignty* if every alternative could potentially be chosen (for example if everyone ranked it first).[1] Note that a dictatorial social choice function is strategy-proof and respects citizen sovereignty: whatever the "dictator" puts first is chosen, and neither the "dictator" nor any other individual can benefit by reporting something other than their true preferences. The Gibbard-Satterthwaite Theorem establishes that the *only* social choice functions that are strategy-proof and respect citizen sovereignty are dictatorial.

A *Pareto efficient* social choice function yields a particular alternative in any profile where that alternative is ranked first by every individual. A *monotonic* social choice function satisfies the condition that if an alternative is chosen in one profile then it must also be chosen in another profile in which the chosen alternative has not been demoted relative to any other alternative by any individual; put differently, if a is chosen in one profile but not a second profile, then monotonicity implies that there is an individual who ranked a above some another alternative b in the first profile but not the second.

In what follows we will use the following result due to Muller and Satterthwaite (1977):

[1] This property is also referred to as *onto* (e.g. Reny (2001)).

Lemma 1 *If a social choice function is strategy-proof and respects citizen sovereignty, it is Pareto efficient and monotonic.*

Proof (Reny, 2001) Suppose that social choice function f yields alternative a for a given profile. Now suppose that individual i alters her preference ordering while ensuring that a is not demoted relative to any other alternative: for example, perhaps she randomly reshuffles the alternatives above a. What alternative is chosen by f for the altered profile that includes i's reshuffled preferences? Because f is strategy-proof, the social choice for the altered profile could only be some $b \neq a$ if i ranked b above a in the altered profile but not in the original profile, but we required that a not be demoted relative to any alternative in the reshuffling, so the social choice must remain a. If a second individual now reshuffled her preferences in the same way, the social choice would have to remain a for the same reason. The implication is that if the social choice is a for one profile, it must remain a in any profile in which no one allowed a to be demoted relative to any other alternative. Thus strategy-proofness implies monotonicity.

Now suppose that a is chosen for some profile. Monotonicity implies that a must also be chosen in another profile where every individual puts a at the top of their preference ordering. Citizen sovereignty requires that every alternative be chosen for some profile. Thus strategy-proofness plus citizen sovereignty implies Pareto efficiency. ∎

Lemma 1 helps clarify what is problematic about the seemingly-desirable property of strategy-proofness: it requires the same choice to be made in profiles where preferences differ quite substantially. For example, strategy-proofness requires that if alternative a is chosen in one profile, it must also be chosen in another profile that is the same except that some other alternative b is ranked first by everyone who ranked b anywhere above a in the first profile. Monotonicity captures this unresponsiveness to certain large changes in preferences, and the proof in the next section shows that this problematic feature of strategy-proofness is the kernel of Gibbard and Satterthwaite's negative result.

2. PROOF

We prove the Gibbard-Satterthwaite Theorem by first proving a simple lemma:

Lemma 2 *Consider a profile (with three or more alternatives) in which an alternative a is ranked last by $n_l \geq 0$ individuals and first by one individual, and in which some other alternative c is ranked first by the remaining $n_r \geq 0$ individuals:*

$$\begin{array}{ccc} (n_l) & (1) & (n_r) \\ \cdot & a & c \\ \vdots & \vdots & \vdots \\ a & \cdot & \cdot \end{array} \tag{1}$$

A strategy-proof social choice function that chooses a for this profile and respects citizen sovereignty must be dictatorial.

Proof Call the one individual who ranks a first in profile 1 the "pivotal individual"; call the n_l individuals who rank a last in profile 1 the "left group" and the n_r individuals who rank c first the "right group". Consider another profile exactly like profile 1 except that c is moved to the top for the left group and just below a for the pivotal individual:

$$\begin{array}{ccc} (n_l) & (1) & (n_r) \\ c & a & c \\ \cdot & c & \cdot \\ \vdots & \vdots & \vdots \\ a & \cdot & \cdot \end{array} \tag{2}$$

By monotonicity (required along with Pareto efficiency by Lemma 1), the social choice must still be a, as no alternative has changed in ranking relative to a for any individual. Now consider another profile exactly like profile 2 except that a is last for the right group:

$$\begin{array}{ccc} (n_l) & (1) & (n_r) \\ c & a & c \\ \cdot & c & \cdot \\ \vdots & \vdots & \vdots \\ a & \cdot & a \end{array} \tag{3}$$

Every alternative apart from a is below c for every individual, so by monotonicity and Pareto the social choice must either be a or c.[2] But monotonicity also indicates that the social choice must not be c; if it were, c must also be chosen in profile 2. Therefore the social choice for profile 3 must be a. But this implies a dictatorship of the pivotal individual. To see this, note that (by monotonicity) the social choice must also be a for *any* profile in which a appears first in the pivotal individual's preference ordering. Thus a must also be chosen in profile 4.i below; a may not, however, be chosen in profile 4.ii (because b is higher than a in everyone's rank order) nor can c be chosen (by monotonicity with respect to profile 4.i), so b must be chosen in profile 4.ii, and by monotonicity b must also be chosen in profile 4.iii, in which a is moved below c for the pivotal individual. But the choice of b in profile 4.iii implies the choice of b in profile 4.iv, in which b is moved to the bottom for everyone but the pivotal individual: if c were chosen it must also be chosen in profile 4.iii by monotonicity, and a is below c for all individuals.

$$\begin{array}{ccc} & \text{4.i} & \\ (n_l) & (1) & (n_r) \\ c & a & c \\ b & b & b \\ \cdot & c & \cdot \\ \vdots & \vdots & \vdots \\ a & \cdot & a \end{array} \qquad \begin{array}{ccc} & \text{4.ii} & \\ (n_l) & (1) & (n_r) \\ c & b & c \\ b & a & b \\ \cdot & c & \cdot \\ \vdots & \vdots & \vdots \\ a & \cdot & a \end{array}$$

[2] If the social choice were some other alternative x, then by monotonicity it would have to remain x if c were moved above a for the pivotal individual, but this would violate Pareto.

$$\begin{array}{ccc} & \text{4.iii} & \\ (n_l) & (1) & (n_r) \\ c & b & c \\ b & c & b \\ \cdot & a & \cdot \\ \vdots & \vdots & \vdots \\ a & \cdot & a \end{array} \qquad\qquad \begin{array}{ccc} & \text{4.iv} & \\ (n_l) & (1) & (n_r) \\ c & b & c \\ \cdot & c & \cdot \\ \vdots & \vdots & \vdots \\ b & \cdot & b \end{array}$$

What is true for alternative b would be true for any alternative, so the choice of a in profile 1 implies that the pivotal individual's first choice is the social choice for all profiles. Note finally that the argument above holds if $n_l = 0$ or $n_r = 0$ (or both, the trivial case in which there is only one individual). $\blacksquare$

Lemma 2 suggests a simple proof of the Gibbard-Satterthwaite Theorem for the special case of three individuals and three alternatives:

Proposition 1 *Given three individuals and three alternatives, any strategy-proof social choice function that respects citizen sovereignty is a dictatorship.*

Proof A social choice function for three individuals and three alternatives must yield a social choice for this profile, sometimes called a Condorcet triple:

$$\begin{array}{ccc} (1) & (1) & (1) \\ a & b & c \\ b & c & a \\ c & a & b \end{array} \tag{4}$$

The conditions for Lemma 2 are satisfied for all three alternatives. $\blacksquare$

We can also use Lemma 2 along with the approach of Geanakoplos (2005) (see also Reny, 2001) to prove the more general Gibbard-Satterthwaite Theorem:

Proposition 2 *Given three or more alternatives, any strategy-proof social choice function that respects citizen sovereignty is a dictatorship.*

Proof Starting with a profile in which all individuals put alternative a first, b second, and z last, consider a sequence of profiles obtained by choosing one individual and raising z in that individual's preference ordering one rank at a time until it reaches the top, then choosing the next individual and doing the same, etc., until all individuals put z first, a second, and b third. By Pareto (required along with monotonicity by Lemma 1), at some point in the sequence the social choice must change from a to z. Let profile 5 indicate the profile just before this occurs (and thus a is chosen) and profile 6 indicate the profile just after this occurs (and thus z is chosen):

$$\begin{array}{ccc} (n_l) & (1) & (n_r) \\ z & a & a \\ a & z & b \\ \vdots & \vdots & \vdots \\ \cdot & \cdot & z \end{array} \rightarrow a \tag{5}$$

$$\begin{array}{ccc} (n_l) & (1) & (n_r) \\ z & z & a \\ a & a & b \\ \vdots & \vdots & \vdots \\ \cdot & \cdot & z \end{array} \rightarrow z \tag{6}$$

n_l indicates the size of the "left group", i.e. those who put z first just before the social choice changes to z, and n_r indicates the size of the "right group", i.e. those who put z last just before the social choice changes to z; the "pivotal individual" is the individual who changes the social choice from a to z by moving z above a. The choice of z in profile 6 indicates that, by monotonicity, z must also be chosen in profile 7 (which is obtained by moving a down for the left group and the right group):

$$\begin{array}{ccc} (n_l) & (1) & (n_r) \\ z & z & b \\ \cdot & a & \cdot \\ \vdots & \vdots & \vdots \\ \cdot & \cdot & a \\ a & \cdot & z \end{array} \tag{7}$$

Now, consider this profile, which is obtained from profile 7 by switching z and a for the pivotal individual:

$$\begin{array}{ccc} (n_l) & (1) & (n_r) \\ z & a & b \\ \cdot & z & \cdot \\ \vdots & \vdots & \vdots \\ \cdot & \cdot & a \\ a & \cdot & z \end{array} \tag{8}$$

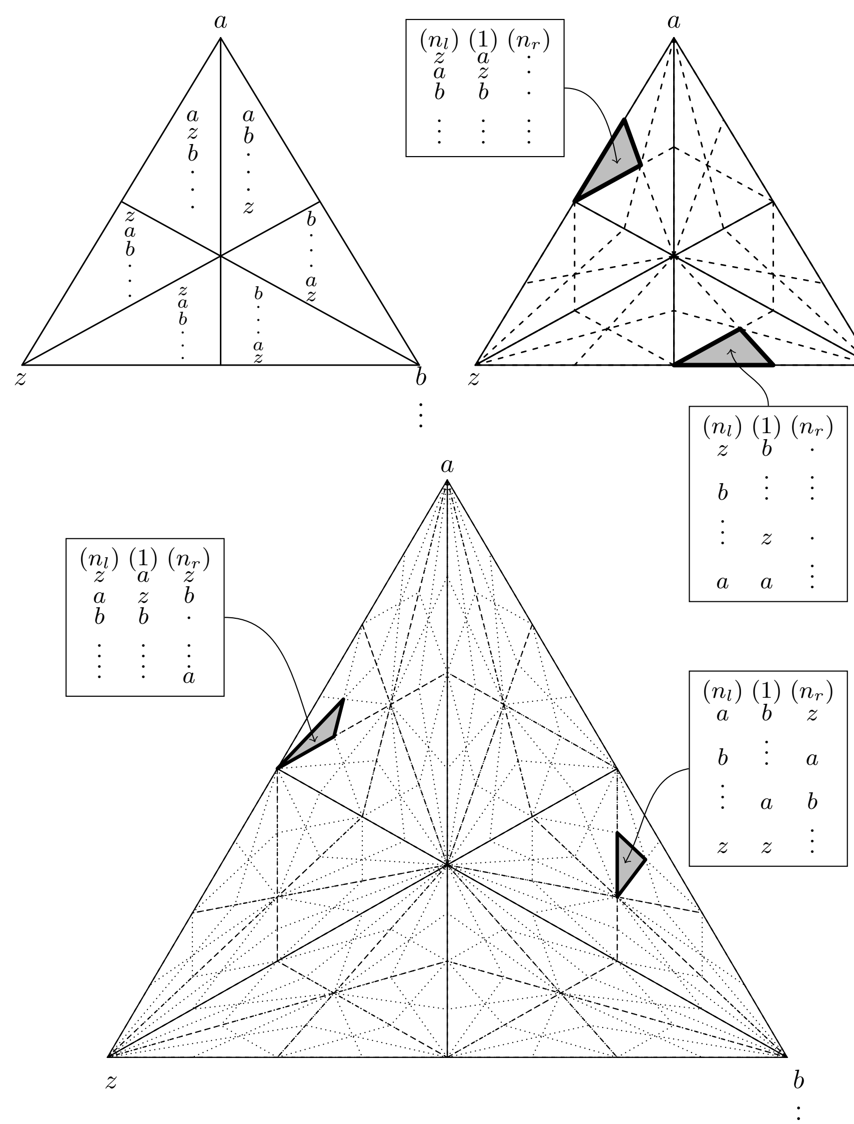
By monotonicity the social choice must be z or a; if any other alternative were chosen it would have to be chosen in Profile 7. But the social choice cannot be z, because otherwise monotonicity would require that z be chosen in profile 5. Thus the social choice is a. Profile 8 meets the conditions of Lemma 2, so the social choice function is a dictatorship. ∎

3. GRAPHICAL EXPLICATION

Figure 2 provides a graphical account of the proof. It relies on what I call a "nested ternary plot" for depicting preference profiles in two-dimensional space. Figure 1 explains the nested ternary plot. We start (upper left) with a triangle with vertices labeled a, z, and $b \cdots$ and divide it into six sub-triangles, each of which corresponds to a unique way to order alternatives a, z, and $b \cdots$ (where $b \cdots$ is a bundle of alternatives headed by b).[3]

[3] For a given sub-triangle the first alternative in the preference ordering is given by the label at the vertex that the triangle touches, and the second alternative is given by the closer

Figure 1: The nested ternary plot



NOTE: In the top left plot, we show the six possible orderings of a, z, and the collection of alternatives headed by b; the ordering in each sub-triangle reflects the distance to the vertices of the large triangle. In the top right plot we divide each of the small triangles in the same way to add the preference orderings of the "left group"; again the ordering reflects distances to the vertices of the large triangle. In the bottom plot we again divide each of the small triangles to add the preference orderings of the "right group". Profiles are given for the highlighted triangles.

Next (top right), within each of the six solid-line sub-triangles we record the preferences of the "left group" in the same way: we divide each triangle into six, and assign the preferences of the "left group" in the same pattern established in the top left plot (middle left triangle indicates z first and a second, etc.). Two of the resulting triangles are highlighted in the figure, with the corresponding preference orderings for the left group and the pivotal individual indicated.

Finally (bottom plot), we apply the same procedure to record the preferences of the right group within each dashed-line triangle. Each of the $6^3 = 216$ small triangles in this plot represents a unique combination of preferences over a, z, and $b \cdots$ for the pivotal individual (solid triangles), left group (dashed lines), and right group (dotted lines). Again two of the profiles are labeled. Triangles near one of the vertices of the large triangle indicate unanimity or near-unanimity for the alternative labeled at that vertex; triangles near the center of the triangle indicate high levels of disagreement. Two triangles that share a dotted-line side indicate profiles that differ only in one binary relation for the right group; otherwise, proximity is only a rough guide to similarity because of the nested nature of the representation.

Now we turn to Figure 2, which uses the nested ternary plot to provide a graphical account of the proof presented above. The symmetrical shaded pattern indicates profiles that satisfy the conditions of Lemma 2: if the social choice function yields the pivotal individual's top choice (and is strategy-proof and respects citizen sovereignty) it must be dictatorial. In the sequence of profiles considered in the proof there must be a pair of profiles 5 and 6 (labeled on the figure) where the social choice changes from a to z as the pivotal individual moves z above a. Neither of these choices directly implies dictatorship, but the choice of z in profile 6 implies the choice of z in profile 7, and together with the choice of a in profile 5 this implies the choice of z in profile 8, which satisfies the conditions of Lemma 2 and thus implies dictatorship.

4. DISCUSSION

How do we know that a non-dictatorial strategy-proof social choice function does not exist? The approach in this paper is to first show that a strategy-proof social choice function must be a dictatorship if it chooses a particular alternative in a profile with certain characteristics (Lemma 2) and then to show that every strategy-proof social choice function must make such a choice.

The Gibbard-Satterthwaite Theorem is surprising only if one fails to recognize that strategy-proofness implies not just imperviousness to manipulation but also imperviousness to certain changes in (sincere) preferences across profiles. What is attractive about strategy-proofness is that the social choice does not change when an individual strategically moves an alternative up in her ranking. But because the social choice function cannot distinguish true preferences from strategic misrepresentations, strategy-proofness also of course implies that the social choice must not change in certain circumstances where one

of the other vertices. These orderings correspond to the preferences of the pivotal individual.

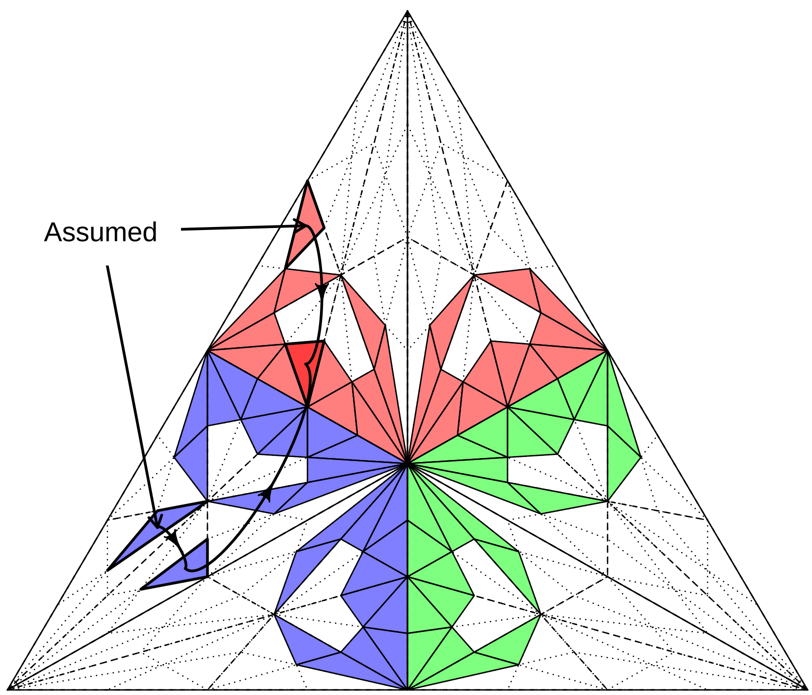


Figure 2: **Graphical account of proof**: The symmetrical shaded pattern indicates profiles that satisfy the conditions of Lemma 2: if the pivotal individual's first choice is chosen, the social choice function must be a dictatorship. The straight arrows point to profiles 5 and 6, i.e. the profiles just before and after the pivotal individual reverses the order of z and a in her preference ordering. The curved lines indicate the implications of these social choices – first, the social choice of the pivotal individual's top-ranked alternative in profile 7 (toward the bottom-left corner), and then the social choice of the pivotal individual's top-ranked alternative in profile 8, which is in the set of profiles satisfying the conditions of Lemma 2.

or more individuals *sincerely* move an alternative up in their rankings. Put simply, strategy-proofness implies preference-proofness in important respects; fundamentally, this is why there are no appealing strategy-proof social choice functions.

Because the Gibbard-Satterthwaite Theorem and Arrow's Theorem can be proved in essentially the same way (Reny, 2001; Eliaz, 2004), intuition that we gain about the first theorem can build our understanding of the second. In particular, the monotonicity property required of choices in the Gibbard-Satterthwaite Theorem is analogous to the IIA property required of rankings in Arrow's Theorem: both properties require the output (choice or ranking) to remain the same even as profiles change substantially.[4] Arrow's Theorem can be proven (e.g. Reny, 2001; Geanakoplos, 2005) in the same manner as shown here: starting from a situation in which one individual is decisive, it emerges that if IIA holds, along with Pareto efficiency, this individual must be decisive in *all* situations.

REFERENCES

Balinski, Michel L and Rida Laraki. 2010. *Majority judgment: measuring, ranking, and electing.* MIT press.

Barberá, Salvador. 1983. "Strategy-proofness and pivotal voters: a direct proof of the Gibbard-Satterthwaite theorem." *International Economic Review* pp. 413–417.

Dummett, Michael and Robin Farquharson. 1961. "Stability in voting." *Econometrica* pp. 33–43.

Eliaz, Kfir. 2004. "Social aggregators." *Social Choice and Welfare* 22(2):317–330.

Geanakoplos, John. 2005. "Three brief proofs of Arrows impossibility theorem." *Economic Theory* 26(1):211–215.

Gibbard, Allan. 1973. "Manipulation of voting schemes: a general result." *Econometrica* pp. 587–601.

Mueller, Dennis C. 2003. *Public choice III.* Cambridge University Press.

Muller, Eitan and Mark A Satterthwaite. 1977. "The equivalence of strong positive association and strategy-proofness." *Journal of Economic Theory* 14(2):412–418.

[4] Another way in which they are analogous is that both properties may not seem indispensable, but they are implied by intuitive properties that seem attractive: monotonicity is implied by strategy-proofness (as shown in Lemma 1 above), and IIA is implied by the independence of social rankings between two alternatives to the presence or absence of another alternative in the choice set (Balinski and Laraki, 2010, pg. 59).

Reny, Philip J. 2001. "Arrows theorem and the Gibbard-Satterthwaite theorem: a unified approach." *Economics Letters* 70(1):99–105.

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

Shepsle, Kenneth. 2010. *Analyzing Politics: Rationality, Behavior, and Institutions.* 2 ed. W.W. Norton.

Vickrey, William. 1960. "Utility, strategy, and social decision rules." *The Quarterly Journal of Economics* pp. 507–535.