Alternatives vs. Outcomes: A Note on the Gibbard-Satterthwaite Theorem

Tjark Weber
tw333@cam.ac.uk
University of Cambridge
Computer Laboratory
15 JJ Thomson Avenue
Cambridge CB3 0FD
United Kingdom

Abstract

The Gibbard-Satterthwaite theorem is a well-known theorem from the field of social choice theory. It states that every voting scheme with at least 3 possible outcomes is dictatorial or manipulable. Later work on the Gibbard-Satterthwaite theorem frequently does not distinguish between alternatives and outcomes, thereby leading to a less general statement that requires the voting scheme to be onto. We show how the Gibbard-Satterthwaite theorem can be derived from the seemingly less general formulation.

JEL classification: D71

Keywords: Gibbard-Satterthwaite theorem; infeasible alternatives

1 Introduction

The Gibbard-Satterthwaite theorem [Gib73, Sat75] states that every voting scheme with at least 3 possible outcomes must be dictatorial or manipulable. The importance of the theorem has been widely recognized [DS00, Tay05, FS06]. Arguably, any practically useful voting scheme should be both non-manipulable and non-dictatorial. The Gibbard-Satterthwaite theorem, however, shows that such a scheme is mathematically impossible, provided there are at least 3 possible outcomes. Following the seminal work of Gibbard

and Satterthwaite, a number of alternative proofs of the theorem have been published [Gär77, SS78, Bar83, Sve99, Ben00, Ren01].

Recent work on the Gibbard-Satterthwaite theorem frequently does not distinguish between the set of *alternatives*, i.e., the set of things to choose from, and the set of possible *outcomes*, i.e., the range of the voting scheme. This leads to a less general formulation of the theorem, which explicitly requires alternatives and possible outcomes to be the same. Take [Sve99], for instance, where the Gibbard-Satterthwaite theorem is stated as follows:

> "A strategy-proof voting rule that is onto is dictatorial if the number of alternatives is at least three."

This is also the statement established by Nipkow [Nip09], who presents computer-checked proofs of the Gibbard-Satterthwaite theorem and of Arrow's impossibility theorem [Arr50] in higher-order logic. Likewise, Reny [Ren01] (using a definition of Pareto efficiency that immediately implies surjectivity) proves a similarly weakened version of the Muller-Satterthwaite theorem [MS77]:

> "If $\#A \geq 3$ and $f\colon \mathcal{L}^N \to A$ is Pareto efficient and monotonic, then f is a dictatorial social choice function."

Also witness the Wikipedia on-line encyclopedia [Wik09], which gives further evidence of the less general formulation's widespread use:

> "The Gibbard-Satterthwaite theorem [...] states that, for three or more candidates, one of the following three things must hold for every voting rule:
>
> 1. The rule is dictatorial [...], or
> 2. There is some candidate who cannot win, under the rule, in any circumstances, or
> 3. The rule is susceptible to tactical voting [...]."

Note the second condition, which suggests that the theorem otherwise only applies to voting schemes that are onto.

Given this formulation of the Gibbard-Satterthwaite theorem, the uninitiated may be thrown off by the possibility to extend the set of candidates with one or more infeasible "dummy" alternatives. The less general formulation then seems to permit the existence of a non-dictatorial, non-manipulable voting scheme whose range is precisely the set of feasible alternatives. Voters would be asked to rank the dummies together with the real alternatives, but this could be considered a minor nuisance.

From the original formulation of the Gibbard-Satterthwaite theorem, we can see directly that this prestidigitation must be futile. In this paper, we show how the original theorem can be recovered from the (seemingly) less general formulation that is predominant in the recent literature. While this result should be well-known to old stagers, the proof is at least not completely trivial: if we start with a voting scheme that is not onto and simply restrict its codomain (i.e., the set of alternatives) to the set of possible outcomes, the resulting function, while onto, is generally not a voting scheme anymore. A voting scheme's domain and codomain are interrelated; votes rank every alternative, not just those which are feasible. The following Section 2 introduces some notation and basic definitions, while Section 3 gives the proof.

2 Basic Definitions

Let A denote a finite set of *alternatives*, and let $\mathcal{L}$ denote the set of strict linear orders, or (strict) *rankings*, on A.[1] Fix a positive integer N. The set of *individuals* then is $\{1, \ldots, N\}$. A function $f\colon \mathcal{L}^N \to A$ is called a *voting scheme*. Elements in $R := \mathsf{range}(f)$ are called *(possible) outcomes* of the scheme, while elements in $A \setminus R$ are called *infeasible*. As usual, we say that f is *onto* iff $R = A$. We write $L|_R$ for the restriction of a ranking L to elements in R, and $\mathcal{L}|_R$ for the set $\{L|_R \mid L \in \mathcal{L}\}$.

Definition 1 (SP)**.** A voting scheme $f\colon \mathcal{L}^N \to A$ is *strategy-proof* iff

$$f(L_1, \ldots, L'_i, \ldots, L_N) \leq_{L_i} f(L_1, \ldots, L_i, \ldots, L_N)$$

for every individual i (i.e., casting a vote L'_i that is perhaps different from i's sincere ranking L_i will not improve the outcome, as measured by L_i). A voting scheme is *manipulable* iff it is not strategy-proof.

Definition 2. A voting scheme $f\colon \mathcal{L}^N \to A$ is *dictatorial* iff there exists an individual i such that (for all $x \in R$) $f(L_1, \ldots, L_N) = x$ if and only if x is at the top of i's restricted ranking $L_i|_R$.

The Gibbard-Satterthwaite theorem, as stated and proved in [Gib73], then reads as follows.

[1] Strictness is a common assumption in the literature on the Gibbard-Satterthwaite theorem, going back to Satterthwaite himself [Sat75], but made merely to simplify the presentation. Our results can easily be extended to non-strict rankings, i.e., rankings which allow ties.

Theorem 3 (Gibbard-Satterthwaite)**.** *Every voting scheme with at least three outcomes is either dictatorial or manipulable.*

Another desirable and well-known property of voting schemes, Arrow's independence of irrelevant alternatives [Arr50], will also be useful to establish our main result.

Definition 4 (IIA)**.** A voting scheme $f\colon \mathcal{L}^N \to A$ is *independent of irrelevant alternatives* iff $f(L_1, \ldots, L_N) = f(L'_1, \ldots, L'_N)$ whenever $L_i|_R = L'_i|_R$ for every individual i.

3 The Proof

We take the less general formulation of the Gibbard-Satterthwaite theorem from [Sve99], slightly reworded to make the difference to Theorem 3 more obvious.

Theorem 5. *Every voting scheme with at least three alternatives that is onto is either dictatorial or manipulable.*

Theorem 5 is an immediate corollary of Theorem 3: every voting scheme with at least three alternatives that is onto is, quite obviously, a voting scheme with at least three outcomes, hence dictatorial or manipulable by Theorem 3.

We now show that Theorems 3 and 5 are in fact equivalent, i.e., that the original Gibbard-Satterthwaite theorem can be derived from the (seemingly) less general formulation. We prove a lemma first.

Lemma 6 (SP implies IIA)**.** *Every strategy-proof voting scheme $f\colon \mathcal{L}^N \to A$ is independent of irrelevant alternatives.*

This lemma is known. See [MS77, p. 414, footnote 2], for instance, where strategy-proofness is shown to be equivalent to a condition called strong positive association (SPA for short): "In addition, one can show that if, in our definition of a voting procedure, we permitted the set [...] of feasible alternatives to vary over some universal set of alternatives, then SPA also implies Arrow's independence of irrelevant alternatives condition." We supply the proof of the lemma, which was omitted in Muller's and Satterthwaite's paper.

Proof. It suffices to show $f(L_1, \ldots, L_i, \ldots, L_N) = f(L_1, \ldots, L'_i, \ldots, L_N)$, provided $L_i|_R = L'_i|_R$, where i is an arbitrary (but fixed) individual. The more general lemma then follows by a straightforward induction on N, using transitivity of equality.

Since f is strategy-proof, we have

$$f(L_1, \ldots, L'_i, \ldots, L_N) \leq_{L_i} f(L_1, \ldots, L_i, \ldots L_N),$$

and likewise

$$f(L_1, \ldots, L_i, \ldots, L_N) \leq_{L'_i} f(L_1, \ldots, L'_i, \ldots L_N).$$

Both sides of these inequations are in R, so the assumption $L_i|_R = L'_i|_R$ yields

$$f(L_1, \ldots, L_i, \ldots L_N) \leq_{L_i} f(L_1, \ldots, L'_i, \ldots L_N) \leq_{L_i} f(L_1, \ldots, L_i, \ldots L_N).$$

Hence $f(L_1, \ldots, L_i, \ldots, L_N) = f(L_1, \ldots, L'_i, \ldots, L_N)$ by irreflexivity and transitivity of $<_{L_i}$. □

Lemma 6 shows that the outcome of a strategy-proof voting scheme is independent of the ranking of infeasible alternatives. While this seems a desirable property, unfortunately it also means that we do not gain any additional freedom in the scheme by asking voters to provide a ranking for dummies. This directly leads us to our main result, a proof of the Gibbard-Satterthwaite theorem from Theorem 5.

Proof. Assume that $f\colon \mathcal{L}^N \to A$ is strategy-proof, with $|R| \geq 3$ (where $R := \mathsf{range}(f)$).

Let $\hat{\cdot}\colon \mathcal{L}|_R \to \mathcal{L}$ be an arbitrary embedding that adds the infeasible alternatives to a ranking of the possible outcomes (e.g., by placing them at the bottom of the ranking; however, any embedding—i.e., any function $\hat{\cdot}$ that satisfies $\hat{L}|_R = L$ for every restricted ranking $L \in \mathcal{L}|_R$—will do).

Define a voting scheme $\hat{f}\colon {\mathcal{L}|_R}^N \to R$ by

$$\hat{f}(L_1, \ldots, L_N) := f(\hat{L}_1, \ldots, \hat{L}_N).$$

It is easy to see that $\hat{f}$ is strategy-proof:

$$\begin{aligned} &\hat{f}(L_1, \ldots, L'_i, \ldots, L_N) = f(\hat{L}_1, \ldots, \hat{L}'_i, \ldots, \hat{L}_N) \\ \leq_{\hat{L}_i} &f(\hat{L}_1, \ldots, \hat{L}_i, \ldots, \hat{L}_N) = \hat{f}(L_1, \ldots, L_i, \ldots, L_N) \end{aligned}$$

since f is strategy-proof, and because both sides of the inequation are in R,

$$\hat{f}(L_1, \ldots, L'_i, \ldots, L_N) \leq_{L_i} \hat{f}(L_1, \ldots, L_i, \ldots, L_N)$$

follows with $\hat{L}_i|_R = L_i$.

Furthermore, $\hat{f}$ is onto: let $x \in R$. Because $R = \mathsf{range}(f)$, we can find L_1, …, $L_N \in \mathcal{L}$ with $f(L_1, \ldots, L_N) = x$. Note that $\widehat{L_i|_R}|_R = L_i|_R$ for each individual i. Thus $\hat{f}(L_1|_R, \ldots, L_N|_R) = f(\widehat{L_1|_R}, \ldots, \widehat{L_N|_R}) = f(L_1, \ldots, L_N) = x$ by Lemma 6, applied to f.

Therefore $\hat{f}$ is dictatorial by Theorem 5, i.e., there exists an individual i such that (for all $x \in R$) $\hat{f}(L_1, \ldots, L_N) = x$ if and only if x is at the top of i's ranking L_i (where $L_1, \ldots, L_N \in \mathcal{L}|_R$).

Now let $L_1, \ldots, L_N \in \mathcal{L}$. Again (using Lemma 6 just like above) we have $f(L_1, \ldots, L_N) = \hat{f}(L_1|_R, \ldots, L_N|_R)$. Thus $f(L_1, \ldots, L_N) = x$ if and only if x is at the top of i's restricted ranking $L_i|_R$. This proves that individual i is a dictator for f. □

The key idea of the above proof is to use the given voting scheme f to define a related scheme $\hat{f}$ to which Theorem 5 can be applied. In particular, $\hat{f}$ must be onto. To achieve this, we consider alternatives in the range of f only, and define $\hat{f}$ for rankings restricted to these alternatives. Independence of irrelevant alternatives then proves that $\hat{f}$ is onto, and that $\hat{f}$ being dictatorial carries over to the original voting scheme f. This is closely related to the *condition of partitioned information* (CPI), stating that the outcome of a voting scheme restricted to a subset of the alternatives depends on rankings over that subset only, which was shown to be equivalent to IIA by Ray [Ray73, Theorem 3].

References

[Arr50] Kenneth J. Arrow. A difficulty in the concept of social welfare. *Journal of Political Economy*, 58(4):328–346, August 1950.

[Bar83] Salvador Barberà. Strategy-proofness and pivotal voters: A direct proof of the Gibbard-Satterthwaite theorem. *International Economic Review*, 24(2):413–417, 1983.

[Ben00] Jean-Pierre Benoît. The Gibbard-Satterthwaite theorem: a simple proof. *Economics Letters*, 69(3):319–322, December 2000.

[DS00] John Duggan and Thomas Schwartz. Strategic manipulability without resoluteness or shared beliefs: Gibbard-Satterthwaite generalized. *Social Choice and Welfare*, 17(1):85–93, January 2000.

[FS06] Allan M. Feldman and Roberto Serrano. *Welfare Economics and Social Choice Theory*. Birkhäuser, 2006.

[Gär77] Peter Gärdenfors. A concise proof of a theorem on manipulation of social choice functions. *Public Choice*, 32:137–142, 1977.

[Gib73] Allan Gibbard. Manipulation of voting schemes: a general result. *Econometrica*, 41(4):587–601, July 1973. Reprinted in Charles K. Rowley, ed., *Social Choice Theory* (Cheltenham: Edward Elgar, 1993).

[MS77] Eitan Muller and Mark A. Satterthwaite. The equivalence of strong positive association and strategy-proofness. *Journal of Economic Theory*, 14(2):412–418, April 1977.

[Nip09] Tobias Nipkow. Social choice theory in HOL: Arrow and Gibbard-Satterthwaite. *Journal of Automated Reasoning*, 2009. To appear.

[Ray73] Paramesh Ray. Independence of irrelevant alternatives. *Econometrica*, 41(5):987–991, September 1973.

[Ren01] Philip J. Reny. Arrow's theorem and the Gibbard-Satterthwaite theorem: a unified approach. *Economics Letters*, 70(1):99–105, January 2001.

[Sat75] Mark A. Satterthwaite. Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, 10:187–217, April 1975.

[SS78] D. Schmeidler and H. Sonnenschein. Two proofs of the Gibbard-Satterthwaite theorem on the possibility of a strategy-proof social choice function. In H. Gottinger and W. Leinfellner, editors, *Decision Theory and Social Ethics: Issues in Social Choice*, pages 227–234. D. Reidel Publishing Company, Dordrecht, 1978.

[Sve99] Lars-Gunnar Svensson. The proof of the Gibbard-Satterthwaite theorem revisited. *Working Papers from Lund University, Department of Economics*, 1, 1999.

[Tay05] Alan D. Taylor. *Social Choice and the Mathematics of Manipulation*. Outlooks. Cambridge University Press, 2005.

[Wik09] Wikipedia. Gibbard-Satterthwaite theorem. In *Wikipedia, The Free Encyclopedia*. 2009. Retrieved September 1, 2009, from `http://en.wikipedia.org/w/index.php?title=Gibbard%E2%80%93Satterthwaite_theorem&oldid=292659871`.