Ville Korpela

A Differential Approach to Gibbard-Satterthwaite Theorem

Aboa Centre for Economics

Discussion Paper No. 74

Turku 2012

ISSN 1796-3133

Printed in Uniprint
Turku 2012

Ville Korpela

A Differential Approach to Gibbard-Satterthwaite Theorem

Aboa Centre for Economics
Discussion Paper No. 74
September 2012 (first draft May 2012)

ABSTRACT

Numerous simple proofs of the celebrated Gibbard-Satterthwaite theorem (Gibbard, 1977, Satterthwaite, 1975) has been given in the literature. These are based on a number of different intuitions about the most fundamental reason for the result. In this paper we derive the Gibbard-Satterthwaite theorem once more, this time in a differen-tiable environment using the idea of potential games (Rosenthal, 1973, Monderer and Shapley, 1996). Our proof is very different from those that have been given previously.

JEL Classification: D71; D82

Keywords: Differentiable function; Gibbard-Satterthwaite theorem; Potential game; Strategy-Proofness

Contact information

ville.korpela (at) utu.fi

A Differential Approach to Gibbard-Satterthwaite Theorem

V. Korpela *

Public Choice Research Centre (PCRC)

September 9, 2012

Abstract

Numerous simple proofs of the celebrated Gibbard-Satterthwaite theorem (Gibbard, 1977, Satterthwaite, 1975) has been given in the literature. These are based on a number of different intuitions about the most fundamental reason for the result. In this paper we derive the Gibbard-Satterthwaite theorem once more, this time in a differentiable environment using the idea of potential games (Monderer and Shapley, 1996).

Keywords: Differentiable function; Gibbard-Satterthwaite theorem; Potential game; Strategy-Proofness

JEL Classification: D71; D82

1. INTRODUCTION

While numerous proofs of the Gibbard-Satterthwaite theorem has been given in the literature, they can all nevertheless be classified into few different categories based on the techniques that are used: (1) The theorem can be derived form Arrows impossibility theorem (e.g. Gibbard, 1977), (2) the dictator can be constructed explicitly (e.g. Benoit, 2000), (3) the theorem can be verified for small number of individuals and/or social alternatives, and then generalized to all possible cases using induction (e.g. Sen, 2001)

*E-mail: *ville.korpela@utu.fi*, tel. +358 2 333 6428, address: Turku School of Economics, Rehtorinpellonkatu 3, 20500 Turku, Finland .

or (4) the *method of option sets* can be used (e.g. Barberá and Peleg, 1990). In addition to these, also some indirect approaches has been taken. One noteworthy and very creative is given by Batteau and Blin (1979). In this paper the theorem is derived using the algebraic theory of ultrafilters.

In the present paper we introduce yet another way to prove the Gibbard-Satterthwaite theorem. The central idea is that we work with a differentiable environment, an approach inspired by the work of Laffont and Maskin (1980). The proof itself is based on the theory of potential games. To be more precise, the proof is obtained as a by product of the remarkably general observation that direct revelation game associated with a strategy-proof social choice function is necessarily a pseudo-potential game.[1] In a differentiable environment Gibbard-Satterthwaite theorem follows from this as a corollary with only a minimal amount of additional work.

2. THE DIFFERENTIABLE SOCIAL CHOICE MODEL

Let $N = \{1, \ldots, n\}$ be the set of individuals, and $X \subseteq \mathbb{R}$ the set of social alternatives to be decided upon. We assume that X is an open interval $\big(\underline{x}, \overline{x}\big)$ to avoid all tedious questions about the behavior at the boundaries of a region.[2] Each individual i has a utility function $U_i\big(x, \theta_i\big)$ over X that is indexed by a parameter $\theta_i \in \big(\underline{\theta}_i, \overline{\theta}_i\big) \subseteq \mathbb{R}$. Let $\Theta = \times_{i=1}^{n} \Theta_i = \times_{i=1}^{n} \big(\underline{\theta}_i, \overline{\theta}_i\big) \subseteq \mathbb{R}^n$. A *Social Choice Function* (SCF) is any mapping $f : \Theta \to X$ that associates any state $\theta \in \Theta$ with a unique social alternative $f(\theta) \in X$.

The idea is that we make the following differentiability assumptions throughout the paper.

Assumption D: For each $i \in N$, the utility function $U_i : X \times \Theta_i \to \mathbb{R}$ is differentiable in the entire domain $X \times \Theta_i \subseteq \mathbb{R}^2$. Furthermore, also the SCF $f : \Theta \to X$ is differentiable. □

As usual, θ_{-i} will denote the profile $\big(\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n\big)$, and $\big(\widehat{\theta}_i, \theta_{-i}\big)$ the profile $\big(\theta_1, \ldots, \theta_{i-1}, \widehat{\theta}_i, \theta_{i+1}, \ldots, \theta_n\big)$. A SCF f is called *Strategy-Proof*

[1] See Shipper (2006) to learn more about pseudo-potential games.

[2] Our result still holds if $X = [\underline{x}, \overline{x}]$ and all functions are right differentiable at $\underline{x}$ and left differentiable at $\overline{x}$.

(SP), if for all $i \in N$, $\theta \in \Theta$ and $\widehat{\theta}_i \in \Theta_i$

$$U_i\Big(f\big(\theta_i, \theta_{-i}\big), \theta_i\Big) \geq U_i\Big(f\big(\widehat{\theta}, \theta_{-i}\big), \theta_i\Big). \tag{1}$$

In a differentiable environment Equation (1) implies that (see Laffont and Maskin, 1980)

$$\frac{\partial U_i\big(f\big(\theta_i, \theta_{-i}\big), \theta_i\big)}{\partial x} \times \frac{\partial f\big(\theta_i, \theta_{-i}\big)}{\partial \theta_i} = 0. \tag{2}$$

As this equation gives only a local conditions, it cannot be sufficient unless we make further assumptions about the utility functions. However, to retain the common assumption of unrestricted domain, we do not want to take this road.

3. STRATEGY-PROOFNESS AND POTENTIAL GAMES

Our proof is based on the theory of potential games. Let $\mathit{\Gamma} = \mathit{\Gamma}\big(u^1, \ldots, u^n; Y\big)$ be a game in strategic form, where n is the number of players, $Y = \times_{i=1}^n Y^i$ is the set of strategy profiles and $u^i : Y \to \mathbb{R}$ is the payoff function of player i. Denote $y^{-i} \in \times_{j \neq i} Y^j \equiv Y^{-i}$. A real-valued function $P : Y \to \mathbb{R}$ is a *pseudo-potential* for $\mathit{\Gamma}$, if for all i and for every $y^{-i} \in Y^{-i}$

$$\emptyset \neq \arg\max_{y^i \in Y^i} P\big(y^i, y^{-i}\big) \subseteq \arg\max_{y^i \in Y^i} u^i\big(y^i, y^{-i}\big). \tag{3}$$

We say that $\mathit{\Gamma}$ is a *pseudo-potential game* if it admits a pseudo-potential function. In a pseudo-potential game one best-reply path can be found by maximizing the common pseudo-potential function. Therefore, in a pseudo-potential game a strong form of strategic alignment is taking place below the surface.

Now fix the SCF $f : \Theta \to X$. For each $\theta \in \Theta$, this SCF defines a game in strategic form: Choose $Y = \Theta$ and define $u^i : \Theta \to R$ by setting:

$$u^i\big(\widehat{\theta}\big) = U_i\big(f(\widehat{\theta}), \theta_i\big) \text{ for all } i \in N \text{ and all } \widehat{\theta} \in \Theta.$$

We denote this game by $\mathit{\Gamma}_f(\theta)$. Our proof of the Gibbard-Satterthwaite theorem is based on the following remarkably general observation.

Theorem 1. Suppose that $f : \Theta \to X$ is SP and that D holds. For each $\theta \in \Theta$, the game $\mathit{\Gamma}_f(\theta)$ is a pseudo-potential game. Furthermore, one possible pseudo-potential $P_\theta : \Theta \to \mathbb{R}$ of this game is

$$P_\theta\big(\widehat{\theta}\big) = \sum_{i=1}^{n} \int\limits_{\tilde{\theta}_{-i}\,\in\,\Theta_{-i}} \Big(U_i\big(f(\widehat{\theta}_i, \tilde{\theta}_{-i}), \theta_i\big) - U_i\big(f(\theta_i, \tilde{\theta}_{-i}), \theta_i\big)\Big) d\tilde{\theta}_{-i}. \tag{4}$$

REMARK 1. We simply assume that the integrals in the definition of $P_\theta : \Theta \to \mathbb{R}$ exist. This holds, for example, if there exists $K \in \mathbb{R}_+$ such that $-K \leq U_i\big(x, \theta_i\big) \leq K$ for all $i \in N$, $\theta_i \in \Theta_i$, and $x \in X$. In other words, if all utility functions are uniformly bounded. $\|$

Proof. Fix $\theta \in \Theta$. To prove the claim, we show that P_θ is a pseudo-potential of $\mathit{\Gamma}_f(\theta)$. Denote

$$P_\theta^i(\widehat{\theta}_i) = \int\limits_{\tilde{\theta}_{-i}\,\in\,\Theta_{-i}} \Big(U_i\big(f(\widehat{\theta}_i, \tilde{\theta}_{-i}), \theta_i\big) - U_i\big(f(\theta_i, \tilde{\theta}_{-i}), \theta_i\big)\Big) d\tilde{\theta}_{-i},$$

so that $P_\theta\big(\widehat{\theta}\big) = \sum_{i \in N} P_\theta^i\big(\widehat{\theta}_i\big)$ by definition. Since only one term in this sum depends on the announcement $\widehat{\theta}_i$ of individual i, we have that for all $\widehat{\theta}_{-i} \in \Theta_{-i}$

$$\arg\max_{\widehat{\theta}_i \in \Theta_i} P_\theta\big(\widehat{\theta}_i, \widehat{\theta}_{-i}\big) = \arg\max_{\widehat{\theta}_i \in \Theta_i} P_\theta^i\big(\widehat{\theta}_i\big).$$

Therefore, we only have to show that for all $i \in N$ and all $\widehat{\theta}_{-i} \in \Theta_{-i}$

$$\arg\max_{\widehat{\theta}_i \in \Theta_i} P_\theta^i\big(\widehat{\theta}_i\big) \subseteq \arg\max_{\widehat{\theta}_i \in \Theta_i} U_i\Big(f(\widehat{\theta}_i, \widehat{\theta}_{-i}), \theta_i\Big). \tag{5}$$

Notice first that $U_i\big(f(\widehat{\theta}_i, \tilde{\theta}_{-i}), \theta_i\big) - U_i\big(f(\theta_i, \tilde{\theta}_{-i}), \theta_i\big) \leq 0$ for all $\tilde{\theta}_{-i} \in \Theta_{-i}$ by SP. This implies that the inequality $P_\theta^i\big(\widehat{\theta}_i\big) \leq 0$ must hold for all $\widehat{\theta}_i \in \Theta_i$, since the integral that defines $P_\theta^i\big(\widehat{\theta}_i\big)$ is taken over a non-positive continuous function by D (as differentiability guarantees continuity). Using SP again, we see that $\theta_i \in \arg\max_{\widehat{\theta}_i \in \Theta_i} P_\theta^i\big(\widehat{\theta}_i\big)$, and if also $\theta_i' \neq \theta_i$ is a maximizer of P_θ^i, then $U_i\big(f(\theta_i', \tilde{\theta}_{-i}), \theta_i\big) = U_i\big(f(\theta_i, \tilde{\theta}_{-i}), \theta_i\big)$ for all $\tilde{\theta}_{-i} \in \Theta_{-i}$ in particular.[3] It

[3]The last claim follows again from the fact that $U_i\big(f(\widehat{\theta}_i, \tilde{\theta}_{-i}), \theta_i\big) - U_i\big(f(\theta_i, \tilde{\theta}_{-i}), \theta_i\big)$ is a non-positive and continuous function of $\tilde{\theta}_{-i}$.

is now easy to see that inclusion (5) must hold (a simple application of SP for the third time). ■

Although pseudo-potential game is admittedly one of the weakest notions of a potential game, it will be sufficient for our purposes. This is a blessing since we really cannot hope for more at this general level.

4. THE GIBBARD-SATTERTHWAITE THEOREM IN A DIFFERENTIABLE ENVIRONMENT

Now that we have all this machinery in place, the influential Gibbard-Satterthwaite theorem will follow as a simple corollary. Before we conclude this paper with the main result, we need one last assumption.

Assumption (Comprehensive Domain). For all $x \in X$ and all $i \in N$, there exists $\theta_i^x \in \Theta_i$, such that $\arg\max_{y \in X} U_i\big(y, \theta_i^x\big) = \{x\}$. □

We need this assumption to establish a connection between the parameter $\theta_i \in \Theta_i$ and the utility function $U_i(x, \theta_i)$. It is impossible to get anything out without this kind of link.

Theorem 2 (The Differential Gibbard-Satterthwaite Theorem). Suppose that D hold. If SCF $f : \Theta \to X$ is SP and has a comprehensive domain, then it must be either dictatorial (on the range $f(\Theta)$) or a constant.

Proof. The pseudo-potential $P_\theta : \Theta \to \mathbb{R}$ of Theorem 1 must satisfy

$$\left.\frac{\partial P_\theta\left(\widehat{\theta}_i, \widehat{\theta}_{-i}\right)}{\partial \widehat{\theta}_i}\right|_{\widehat{\theta}_i = \theta_i} = 0 \quad \text{for all } i \in N \text{ and all } \widehat{\theta}_{-i} \in \Theta_{-i}. \tag{6}$$

Lets assume that f is not dictatorial nor constant. As it is, there must then exist $\theta_i' \in \Theta_i$ and $\theta_{-i}', \theta_{-i}'' \in \Theta_{-i}$, such that $f(\theta_i', \theta_{-i}') = x \neq y = f(\theta_i', \theta_{-i}'')$. Equation (6) implies that $P_{(\theta_i', \widehat{\theta}_{-i})}$ is constant on the set $\{(\theta_i', \theta_{-i}) \mid \theta_{-i} \in \Theta_{-i}\}$ for all $\widehat{\theta}_{-i} \in \Theta_{-i}$. This is impossible when $j \neq i$ has the utility function $U_j(\,\cdot\,, \theta_j^x)$. This proves the claim. ■

5. FINAL REMARKS

Most proofs for the Gibbard-Satterthwaite theorem proceed by constructing the dictator. This leaves a lot to hope for if we want to explaine why the dictator must exist in the first place. The question that we should ultimately answer is why there has to exist a dictator rather than a system with shared power. In this paper we have tried to seek a somewhat more fundamental reason than just a direct proof. Our explanation consists of two parts: (1) Strategy-proofness is such a strong form of incentive alignment that it entails the existence of a potential function, a pseudo-potential function to be exact, which represents a very strong form of strategic alignment between the individuals as this function is common to all, and (2) when the domain of preferences is large, or unrestricted, this pseudo-potential function must be constant (from essential parts, not entirely) to produce the requisite strategic alignment. This demonstrates clearly how the Gibbard-Satterthwaite theorem is an interplay between two things, the strong requirement of incentive alignment on the one hand, and the severe lack of information on the other.

References

BARBERÁ, S. and PELEG, P. (1990): "Strategy-Proof Voting Schemes with Continuous Preferences". *Social Choice and Welfare*, **7**: 31-38.

BATTEAU, P. and BLIN, J.-M. (1979): Elements for New Insights into the Gibbard-Satterthwaite Theorem. In *Aggregation and Revelation of Preferences*, edited by J.-J. Laffont, North-Holland, Netherlands.

BENOÎT, J.-P. (2000): "The Gibbard-Satterthwaite Theorem: A Simple Proof". *Economics Letters* **69**: 319-322.

GIBBARD, A. (1977): "Manipulation of Voting Schemes: A General Result". *Econometrica* **41**: 587-600:

LAFFONT, J.-J. and MASKIN, E. (1980): "A Differential Approach to Dominant Strategy Implementation". *Econometrica* **48**: 1057-1520.

MONDERER, D. and SHAPLEY, L. (1996): "Potential Games". *Games and Economic Behavior* **14**: 124-143.

SATTERTHWAITE, M., A. (1975): "Strategy-Proofness and Arrow's Condition: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions". *Journal of Economic Theory* **10**: 187-217.

SCHIPPER, B. (2004): Pseudo-Potential Games. Working Paper, University of Bonn, mimeo.

SEN, A. (2001): "Another Direct Proof of the Gibbard-Satterthwaite Theorem". *Economics Letters* **70**: 381-385.

The **Aboa Centre for Economics (ACE)** is a joint initiative of the economics departments of the Turku School of Economics at the University of Turku and the School of Business and Economics at Åbo Akademi University. ACE was founded in 1998. The aim of the Centre is to coordinate research and education related to economics.

Contact information: Aboa Centre for Economics, Department of Economics, Rehtorinpellonkatu 3, FI-20500 Turku, Finland.

www.ace-economics.fi

ISSN 1796-3133