August 8, 2002

Voluntary Provision of a Public Good and Individual Morality*

Abstract

We examine, both in general games in strategic form and in games of voluntary provision of a public good, some implications of the assumption that individuals' behavior may obey to ethical codes of conduct. The notion of morality considered captures the intuition (often attributed to Kant) that a moral action leads to the best outcome when it is properly *universalized.* We propose a formalization of this idea which generalizes earlier attempts made in this direction in the literature by allowing the players to differ both in their strategy sets and their preferences. We show that it is easy to find examples of games in which no moral behavior of this type exists or where the only existing "Kantian" code of conduct leads to a Pareto-inefficient outcome. We then more specifically examine the issues of existence and Pareto-efficiency of Kantian norms of behavior in problems of voluntary provision of a public good. We find in this context that there is no conflict between morality and Pareto-efficiency since any Kantian norm of behavior is Pareto-efficient. We also prove the existence of a Kantian norm of individual contribution.

Marc Bilodeau
Department of Economics
IUPUI
Indianapolis, IN
USA 46202-5140
mbilodea@iupui.edu

Nicolas Gravel
Université Paul Valéry, Montpellier
and GREQAM-IDEP
Centre de la Vieille Charité
2, rue de la Charité, 13 002 Marseille
France
gravel@ehess.cnrs-mrs.fr

*The first author gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada. This paper was written while the second author was at THEMA, université de Cergy-Pontoise. We thank (without making them responsible for many remaining shortcomings) Françoise Forges, Jean-François Laslier, Pete Rangazas, Al Slivinski, Rich Steinberg, Alain Trannoy, Lise Vesterlund and two anonymous referees for comments, suggestions and/or helpful discussions. We also thank Ann Burford for proofreading the English text.

1 Introduction

It has been argued by several authors in economics (see for example Laffont (1975), Etzioni (1987) and Bordignon (1990)) that promoting individual morality could help improve the outcome in situations where self-interested behavior is inefficient, for example free-riding in the provision of a public good. One notion of morality often mentioned in this context is Kant's categorical imperative: 'Act only on that maxim through which you can at the same time will that it should become a universal law' Kant (1907) (p.88).[1] In a pioneering contribution, Laffont (1975) interprets Kant's categorical imperative as compelling an individual to undertake any action which he would want everyone else to undertake and suggests that universal adoption of such a principle could overcome the free-rider problem and yield a Pareto-efficient outcome in situations involving the production of a public good. Since then, several authors, including Collard (1978), Sugden (1984), DeJasey (1989), Bordignon (1990), Sandler and Posnett (1992), Sandler (1992), Nida-Rumelin (1993), Binmore (1994) (section 2.4.1), Bergstrom (1995), and Roemer (1996) (ch.6) have also discussed Kant's imperative and so-called Kantian behavior. A recent thoughtful examination of this literature can be found in Wolfelsperger (1999).[2]

In this paper, we seek to clarify what outcomes could be expected if individuals felt compelled by an ethical code of conduct to undertake any action that they would want everyone else to undertake (the common interpretation of Kant's categorical imperative). In

[1] Other notions of individual morality have also been considered in the economic literature, e.g. *reciprocity* in Sugden (1984), *matching* in Guttman (1978) and Cornes and Sandler (1984), and *fair share* in Young (1989).

[2] Among other thingsWolfelsperger (1999) argues that the "common Kantism" discussed in the above literature is not a faithful interpretation of Kant's idea. We will nonetheless use the expression "Kantian" to refer to behavior that accords with this common interpretation of Kant's categorical imperative.

particular, we generalize the earlier work of Laffont (1975) and Bordignon (1990) by allowing the individuals to differ, and we investigate the conditions under which Kantian behavior is possible and desirable.

To illustrate what we are trying to do, reconsider Laffont (1975)'s example of a beach where a large number of individuals come to sunbathe, swim, picnic and drink beer. Suppose that garbage bins are spaced along the beach and that all beach-goers have identical preferences for beach cleanliness and the effort of carrying garbage. For anyone, tidying up and carrying trash to a garbage bin is bothersome, but going to a dirty beach is even more annoying. This situation can be modelled as a prisoner's dilemma and the decision problem faced by a typical beach-goer can be depicted informally in the following payoff matrix:

		all others	
		do not litter	litter
i	do not litter	100	0
	litter	101	1

If this game is played only once, littering is a dominant strategy for everyone, and the only equilibrium has everyone littering. We know from the "Folk theorem" that in repeated plays of this game, other outcomes in which everyone pick up their rubbish become possible as long as individuals do not discount the future too much. In practice however, this is not likely to happen with large numbers of anonymous individuals, because of the difficulty in applying selective punishments to those who do not comply. A factor that might help the beach-goers to reach a better outcome is the *ethical environment* in which they make their decisions.

Suppose for instance that a beach-goer is bound by Kant's categorical imperative. To predict how she will behave, we need to determine what *maxim* she would want to erect as a universal law. We define a maxim as a rule that *prohibits* some actions that the individuals could undertake. Alternatively, a maxim can be viewed as a rule that *prescribes* to individuals the actions that it does not prohibit. Two obvious maxims that could be defined in the beach example are "Nobody should litter" and "Everyone must litter." But there are many others. For example "individual i must litter but no one else should" or "individuals whose first names start with the letter A, D, L and Z must litter, those whose first names start with the letter M or N can do whatever they want, and everyone else must not litter" are also maxims. However, intuitively these last two maxims do not seem very "Kantian." We propose two properties that a maxim must satisfy in order to be compatible with the common interpretation of Kant's categorical imperative: it must treat everyone "similarly" and it must prescribe to everyone actions that are best under the hypothesis that everyone plays an equivalent strategy.

To treat everyone similarly, a maxim must prescribe to everyone actions that are in some sense *equivalent*. The reason the two last maxims above do not seem Kantian is because they arbitrarily single out some individuals and prescribe different actions to them. When individuals are game-theoretically identical (they have the same strategy sets and, up to a permutation, the same preferences for the result of the interaction) as in the beach example, most people would adopt Laffont (1975)'s viewpoint of considering that *equivalent* actions can only be *identical* actions. In this example, only two maxims satisfy this view of equivalence: "Nobody should litter" and "everybody must litter." However, when individuals are different it is not always obvious how a maxim could treat them similarly. For example, is littering by someone who is severely physically impaired, and who must for this reason exert extraordinary effort to collect and carry trash, necessarily equivalent to littering by someone who has no physical impairment?

In this paper, we propose a general notion of equivalence between actions through what we call a *system of universalization.* By this, we simply mean a (moral) equivalence relation between the actions (or strategies) of different players. A system of universalization compares any two actions and determines whether they are morally equivalent. It is outside the scope of this paper to inquire about the origin of this system of universalization. Presumably it will be culturally determined, perhaps ingrained in people's world view from the time their parents and teachers responded to similar behavior with similar rewards and punishments. We take the system of universalization as a primitive datum, like preferences.

The second property that a maxim must satisfy is to yield everyone's most preferred outcome if everyone else is constrained to play a morally equivalent strategy. For a maxim to be congruent with the common interpretation of Kant's categorical imperative, an individual must *want* everyone to abide by it. We suppose that a Kantian individual would make the following reasoning: "Of all the maxims that prescribe equivalent actions to everyone, which one do I prefer? This is the one that I want to erect as a universal law." So in the beach example, each individual compares her payoff when everybody litters to the one when nobody litters, finds that her payoff is higher if nobody litters and concludes that the maxim "do not litter" is the one that she would want to erect as a universal law. Therefore she interprets the categorical imperative as commanding her not to litter. When all individuals are identical, all would reach the same conclusion, and the maxim "do not litter" yields the most preferred outcome for everyone. Maxims that satisfy these two properties will be referred to as *Kantian*. As pointed out by Laffont, universal obedience to a Kantian maxim also leads to a Pareto-efficient outcome in this example.

The situation is more complex when individuals are different. Suppose for example that two Kantian individuals are trying to determine how much to contribute to a public good and that the prevailing notion of equivalence is that contributions of the same amount are morally equivalent. So each asks herself how much she would like everyone to contribute. If they have different income and/or preferences, it is entirely possible that the first contributor would prefer to see everyone contributing, say \$10, while the second would prefer the situation where everyone contributes, say \$50. Then the first individual would want to erect contributions of \$10 into a universal law, while the second would want a universal law that commanded everyone to contribute \$50. In this case, a Kantian maxim would not exist with respect to this system of universalization. The unanimity that is required for a Kantian maxim to exist is automatically obtained when individuals are identical, but not when they are different. As we will see below, it is easy to construct games in which no Kantian maxim exists for all nontrivial systems of universalization. It is also easy to find examples of games in which a Kantian maxim prescribes a Pareto-inefficient profile of actions.

This possibility that Kantian behavior could conflict with Pareto-efficiency must be taken into account in any discussion about the role of individual morality in solving conflicts between private and public interest. Since free-riding in the voluntary provision of a public good is the canonical example of this kind of conflict, we explore the existence and Pareto-efficiency of a Kantian maxim for this class of situations. We show in this setting that for systems of universalization that are *tight* (such that no two distinct actions of a single individual are morally equivalent to each other) and differentiable, any Kantian maxim must be Pareto-efficient. Laffont's central intuition is therefore robust to relaxing the assumption of identical preferences and wealth. We also show that in this class of games there will always exist at least one tight and differentiable system of universalization with respect to which a Kantian maxim can be defined. The Kantian maxim that we construct in order to establish this existence result can be described by the aphorism "Contribute your Lindahl equilibrium amount".

The next section contains formal definitions of the concepts and some examples. Section three contains the main results for games of voluntary provision of a public good.

2 A Formal Definition of Individual Morality[3]

We consider games in strategic form $G = \{N, \times_{i=1}^{n} S_i, < V_i(.) >_{i=1}^{n}\}$ in which $N = \{1, .., n\}$ is the finite set of players, S_i is the set of strategies (actions) available to individual i and $V_i : \times_{j=1}^{n} S_j \longrightarrow \mathbb{R}$ is individual i's payoff function. Given a player $i \in N$, we let $S_{-i} = \times_{j \neq i} S_j$ denote the set of all combinations of strategies that other players could adopt and we denote, when convenient, by $(s_i; s_{-i}) \in S_i \times S_{-i}$ the strategy combination in which player i plays s_i and the other players play the combination $s_{-i} \in S_{-i}$. We now define a **system**

[3] **Our terminology, definitions and notation for binary relations is as follows. By a *binary relation* R on a set Ω, we mean a subset of $\Omega \times \Omega$. Following the convention used in economics, we write $x\ R\ y$ instead of $(x, y) \in R$. Given a binary relation R, we define its *symmetric factor* R_S by $x\ R_S\ y \Longleftrightarrow (x\ R\ y) \wedge (y\ R\ x)$. A binary relation R on Ω is *reflexive* if the statement $x\ R\ x$ holds for every x in Ω, is *transitive* if $x\ R\ z$ follows $x\ R\ y$ and $y\ R\ z$ for any x, y, $z \in \Omega$ and is *symmetric* if $R = R_S$. A symmetric, reflexive and transitive binary relation is called an *equivalence relation.* Given an equivalence relation R on Ω, and given some $\omega \in \Omega$, we denote by $E_R(\omega)$ the equivalence class of ω under R defined by $E_R(\omega) = \{\omega' \in \Omega \mid \omega'\ R\ \omega\}$. It is immediate to see that if R is an equivalence relation, one has $E_R(\omega) \neq \emptyset$ for every $\omega \in \Omega$, either $E_R(\omega) = E_R(\omega')$ or $E_R(\omega) \cap E_R(\omega') = \emptyset$ for every ω and $\omega' \in \Omega$ and $\bigcup_{\omega \in \Omega} E_R(\omega) = \Omega$.**

of universalization.

Definition 1 *A* ***system of universalization*** *is an equivalence relation M on $\bigcup_{i \in N} \{i\} \times S_i$ which satisfies the property that $\forall\ i,\ j \in N,\ \forall\ s \in S_i,\ E_M(i,s) \cap \{j\} \times S_j \neq \emptyset$.*

In this definition, the statement $(i,s)\ M\ (j,s')$ means "strategy s for individual i is morally equivalent to strategy s' for individual j". Equivalence relations are by definition reflexive (the action of any individual is morally equivalent to itself), symmetric (if the play of s by i is morally equivalent to the play of s' by j, then the play of s' by j must also be considered equivalent to the play of s by i) and transitive (if s by player i is considered equivalent to s' by player j which is itself considered equivalent to the play of s'' by player h, then the play of s by i must also be considered equivalent to the play of s'' by h). In this definition, the set $E_M(i,s)$ is interpreted as the set of all pairs of individualized actions that are morally equivalent to the play of s by individual i. This set is not empty since by reflexivity, it contains (i,s). We add the requirement that this set must have a non-empty intersection with the strategy set of every other individual to insure that the system of universalization is complete. This rules out the possibility that a player might have "orphaned" strategies that have no moral equivalent in someone else's strategy set.

A system of universalization M can be represented by a list of correspondences between individual strategy sets. As it turns out, it does not take many correspondences to represent a given system of universalization. It just takes n of them as the following lemma establishes:

Lemma 1 *Let M be a system of universalization defined on $\bigcup_{i \in N} \{i\} \times S_i$. Then, given any individual h, there exists n onto correspondences $\Psi_i^h : S_h \to S_i$ (for $i \in N$) such that for every $i,\ j \in N$, for every $s \in S_i,\ s' \in S_j,\ (i,s)\ M\ (j,s') \Leftrightarrow s \in \Psi_i^h(\overline{s})$ for some $\overline{s} \in \Psi_h^j(s')$ (where $\Psi_h^j : S_j \to S_h$ denote the inverse correspondence of Ψ_j^h defined by $s_h \in \Psi_h^j(s_j) \Leftrightarrow s_j \in \Psi_j^h(s_h)$).*

The proofs of all lemmas and propositions are collected in the appendix. Lemma 1 allows us to represent a system of universalization by n correspondences from one individual strategy set *onto* every individuals' strategy sets. The reference individual h from whose strategy set the correspondences originate is arbitrary.

Definition 1 allows for very loose notions of moral equivalence since it does not preclude the possibility that *many* distinct strategies of some individual j could be morally equivalent to each other or to any strategy of individual i. In particular, definition 1 allows the trivial system of universalization that considers every strategy morally equivalent to every other. Quite intuitively, if the system of universalization is too loose, maxims defined with respect to it would not impose significant constraints on individual behavior. For this reason, and for analytical convenience, much of the analysis in this paper will restrict attention to **tight** systems of universalization.

Definition 2 *A system of universalization M on $\bigcup_{i \in N} \{i\} \times S_i$ is **tight** if , for every $i \in N$ and for every $s \in S_i$, it satisfies $\# E_M(i,s) \cap \{j\} \times S_j = 1$ for any $j \in N$.*

In words, a system of universalization is tight if every strategy has one and only one moral equivalent in every other individual's strategy set. This is a significant restriction. In particular, if the strategy sets of different players do not have the same *cardinality* a system of universalization cannot be tight. As formally established in lemma 2 below (the proof is omitted), tight systems of universalization can be represented by n *one-to-one* functions (instead of correspondences as in lemma 1).

Lemma 2 *Let M be a tight system of universalization defined on $\bigcup_{i \in N} \{i\} \times S_i$. Then, given any individual h, there exists n one-to-one functions $\Psi_i^h : S_h \to S_i$ (for $i \in N$) such that for every i, $j \in N$, for every $s \in S_i$, $s' \in S_j$, (i,s) M $(j,s') \Leftrightarrow s = \Psi_i^h(\Psi_h^j(s'))$ (where $\Psi_h^j : S_j \to S_h$ denote the inverse of Ψ_j^h defined by $s_h = \Psi_h^j(s_j) \Leftrightarrow s_j = \Psi_h^j(s_h)$).*

Other properties could be imposed on a system of universalization if we think they are plausible or desirable. For example, we might wish to restrict the class of admissible systems to those that consider identical strategies by identical individuals as morally equivalent. We might also find it plausible and useful to impose a continuity property on the system of universalization if the strategy sets have some topological structure.

We now introduce the notion of a **maxim**.

Definition 3 *A **maxim** is a subset μ of $\times_{i=1}^{n} S_i$*

As mentioned in the introduction, a maxim is a rule that prescribes to players some actions. We formalize this by defining a maxim as a joint restriction on the strategy sets of the players in the game. For example, the maxim "contribute at least 10% of your income" forces the player who obeys to it to contribute at least that share of her income. In many cases, a maxim may rule out all but one strategy for every player, e.g., "contribute exactly 10% of your income". If this is the case, then μ is a singleton and we say that the maxim is **strict**. When a maxim is strict, we will treat it as an element (rather than a subset) of $\times_{i=1}^{n} S_i$. In general, maxims may restrict quite differently the strategies available to different players (e.g., "everyone but Al, Bob and Chris must contribute at least 10% of their income, Al must contribute exactly 15%, Bob must contribute no more than 5%, and Chris can contribute any nonnegative amount."). Given a system M of universalization, we now define the two properties that a maxim must satisfy in order to qualify as "Kantian." The first property is that a maxim must prescribe morally equivalent strategies to everyone.

Property 1 (moral equivalence) $\forall\ s \in \mu,\ (i, s_i)\ M\ (j, s_j)\ \forall\ i,\ j \in N.$

Second, individuals must *will* that it should become a *universal* law. As discussed in the introduction, we interpret this as meaning that the individual will choose the maxim that gives her the highest payoff given her beliefs about how other individuals would play if they were to behave in a morally equivalent way.

Property 2 (**Universalized rationality**) $\forall\ i \in N$, $\forall\ s \in \mu$,

$$\int\limits_{E_M(i,s_i)\backslash\{i\}\times S_i} V_i(s_i; s_{-i})\pi_i^{s_i} ds_{-i} \geq \int\limits_{E_M(i,\overline{s}_i)\backslash\{i\}\times S_i} V_i(\overline{s}_i, \overline{s}_{-i})\pi_i^{\overline{s}_i} d\overline{s}_{-i}$$

$\forall\ \overline{s}_i \in S_i$, where, for every $s \in S_i$, π_i^s is a Borel belief measure on $E_M(i,s)\backslash\{i\} \times S_i$.

For our purposes, it does not matter how the beliefs mentioned in this definition are formed. For example, individuals could hold rational beliefs in the sense of Bernheim (1984). Another possibility could be that, when considering strategy s_i, individual i believes that others would play the Nash equilibrium strategies of the game in strategic form for which the strategy sets are restricted to $E_M(i,s)$ (if of course this game has one). When the system of universalization is tight, the beliefs are trivial.

As mentioned, we consider a maxim to be *Kantian* if there exists a system of universalization with respect to which it satisfies moral equivalence and universalized rationality. For the sake of completeness, we state this definition formally.

Definition 4 *A maxim* $\mu \subseteq \times_{i\in N} S_i$ *is Kantian if there exists a system of universalization* M *with respect to which it satisfies moral equivalence and universalized rationality.*

We shall sometimes say that a system of universalization M *supports* a Kantian maxim if the considered maxim satisfies moral equivalence and universalized rationality with respect to M.

For latter use, we also recall the definition of Pareto-efficiency.

Definition 5 *(**Pareto-Efficiency**) A strategy profile* $(\widehat{s}_1, ..., \widehat{s}_n) \in \times_{i=1}^n S^i$ *is* Pareto-efficient *if and only if, for every* $(s_1, ..., s_n) \in \times_{i=1}^n S^i$*,* $V_i(s_1, ..., s_n) > V_i(\widehat{s}_1, ..., \widehat{s}_n)$ *for some* $i \in N$ *implies* $V_j(\widehat{s}_1, ..., \widehat{s}_n) > V_j(s_1, ..., s_n)$ *for some* $j \in N$*.*

Now that we have defined a system of universalization and the notion of a Kantian maxim, we illustrate two interesting logical possibilities. Example 1 is a game in which no Kantian maxim can be defined for any tight system of universalization. Example 2 is a game in which any Kantian maxim with respect to a tight system of universalization would prescribe a Pareto-inefficient course of actions.

Example 1: Consider the following two-player game and suppose that individuals can only play pure strategies.

		#2	
		C	D
#1	A	(2,0)	(3,1)
	B	(1,3)	(0,2)

There are three possible systems of universalization in this environment. The tight system M_1 says that $A\ M_1\ C$ and $B\ M_1\ D$; the tight system M_2 says that $A\ M_2\ D$ and $B\ M_2\ C$; and the (non-tight) trivial system M_3 considers all strategies to be equivalent. If the prevailing notion of moral equivalence is M_1, then in order for a maxim μ to satisfy universalized rationality, it must be strict and defined by $\mu_1 = A$ and $\mu_2 = D$ but these two strategies are not morally equivalent. Similarly if the prevailing notion of moral equivalence is M_2, then universalized rationality requires $\mu_1 = A$ and $\mu_2 = C$; this also violates moral equivalence. Hence no Kantian maxim can be defined with respect to either M_1 or M_2. Trivially, the maxim $\mu = \{A, B\} \times \{C, D\}$ ("do whatever you want") is Kantian with respect to M_3 if players have rational beliefs.

Example 2: Consider the following modern version of the "battle of the sexes" game and restrict attention to pure strategies:

		man	
		football	ballet
woman	football	(3,2)	(1,1)
	ballet	(0,0)	(2,3)

If we leave aside the trivial system of universalization that consider all strategies to be equivalent, there are two possible tight systems of universalization in this game. One system, M_1, considers identical actions to be morally equivalent (football for man is morally equivalent to football for woman and ballet for man is morally equivalent to ballet for woman). The second system, M_2, considers football for man to be equivalent to ballet for woman and football for woman to be equivalent to ballet for man. There is no Kantian behavior if the prevailing notion of moral equivalence is M_1 since the man's best choice under this system is ballet while the woman's best choice is football. Yet these two strategies are not morally equivalent. There is a Kantian maxim if the prevailing view of moral equivalence is M_2: Both go alone to their most preferred activity. This outcome is not Pareto-efficient however.

These two examples show that neither the existence nor the Pareto-efficiency of norms of behavior satisfying moral equivalence and universalized rationality with respect to a tight system of universalization can be guaranteed in general. In the next section, we examine the problem in the context of games of voluntary provision of a public good.

3 Morality in games of voluntary provision of a public good

3.1 Notation and assumptions

The model considered in this section is well-known and described for instance in Bergstrom, Blume and Varian (1986). There are n individuals and two goods: A private good, the quantity of which is represented by a variable x, and a public good represented by a variable Z. Each individual $i \in N = \{1, ..., n\}$ is endowed with ω_i units of the private good that she can allocate between her private consumption x_i and a contribution z_i to the production of the public good. The total amount of the public good available for consumption is the sum of all individual contributions so that $Z = \sum_{i \in N} z_i$. We denote by $\omega = \sum_{i \in N} \omega_i$ the total wealth of the community. The preferences of every individual i for the public good and the private good are assumed to be represented by a differentiable, strictly quasi-concave and strictly increasing utility function $U_i : \mathbb{R}^2_+ \to \mathbb{R}$.[4] For any $(\overline{Z}, \overline{x}) \in \mathbb{R}^2_+$, we denote by $MRS^i(\overline{Z}, \overline{x})$ the *marginal rate of substitution* of individual i at $(\overline{Z}, \overline{x})$ defined by

$$MRS^i(\overline{Z}, \overline{x}) = \frac{\frac{\partial U_i(\overline{Z}, \overline{x})}{\partial Z}}{\frac{\partial U_i(\overline{Z}, \overline{x})}{\partial x}} \tag{1}$$

In terms of the game theoretic framework of the preceding section, individual i's strategy set $S_i = [0, \omega_i]$ is the set of all contributions to the public good that she can make. Given a combination of contributions $(z_1, ..., z_n)$, the payoff function of individual i is the function $V_i : \times_{j \in N}[0, \omega_j] \to \mathbb{R}$ defined by $V_i(z_1, ..., z_n) = U_i(\sum_{j \in N} z_j, \omega_i - z_i)$.

Let $(N, \times_{i \in N}[0, \omega_i], < V_i(.) >_{i \in N})$ denote a typical Game of Voluntary Provision of a Public Good (GVPPG). We further impose the following relatively innocuous restriction on

---

[4] That is: (i) (strict quasi-concavity) $U_i(\lambda Z + (1-\lambda)Z', \lambda x + (1-\lambda)x') > U_i(Z', x')$ for every $\lambda \in]0, 1[$ and every (Z, x) , $(Z', x') \in \mathbb{R}^2$ such that $(Z, x) \neq (Z', x')$ and $U_i(Z, x) \geq U_i(Z', x')$ and (ii) (strict increasingness) $U_i(Z, x) > U_i(Z', x')$ for every (Z, x), $(Z', x') \in \mathbb{R}^2$ such that $(Z, x) \neq (Z', x')$ and $(Z, x) \geq (Z', x')$.

the class of games considered herein:

Condition 1 :*For every* $i \in N$, $U_i(\omega_i, 0) > U_i(0, \omega)$.

This says that all individuals prefer contributing nothing and not consuming any public good, to contributing all their income toward the public good even if everyone else also contributes all their income. This condition would be satisfied if private consumption includes an essential item such as food while the public good is not essential. Among other things, this assumption implies that the vector of contributions $(\omega_1, ..., \omega_n)$ cannot be Pareto efficient. We denote by $\mathcal{P}$ the class of all GVPPG that satisfy the properties listed above.

We now note two results (the proofs are omitted) that will be used in the proofs of the propositions below. First we note that a Kantian maxim with respect to a tight system of universalization in a game in $\mathcal{P}$ is necessarily strict.

Lemma 3 *Let* $(N, \times_{i \in N}[0, \omega_i], < V_i(.) >_{i \in N}) \in \mathcal{P}$. *Then* $\mu \subseteq \times_{i \in N}[0, \omega_i]$ *satisfies moral equivalence and universalized rationality with respect to a tight system of universalization only if* $\#\mu = 1$

Second, we recall the sufficiency of Samuelson (1954)'s condition for a Pareto-optimum in the class of games considered here.

Lemma 4 *Let* $(z_1, ..., z_n) \in \mathbb{R}^n_+$ *be a vector of contributions satisfying* $z_i \in [0, \omega_i]$ *for every* i *and* $\sum_{i \in N} MRS(\sum_{j \in N} z_j, \omega_i - z_i) = 1$ *for the GVPPG* $(N, \times_{i \in N}[0, \omega_i], < V_i(.) >_{i \in N}) \in \mathcal{P}$. *Then* $(z_1, ..., z_n)$ *is Pareto-efficient.*

3.2 Pareto-efficiency

In his contribution, Laffont (1975) assumed that individuals were identical. This led him to implicitly define a Kantian maxim with respect to a system of universalization that considers contributions of the same amount to be equivalent. Bordignon (1990) discusses the possibility of generalizing Laffont's model by assuming that individuals have different incomes. He therefore proposes two alternative systems of universalization: (i) equal contributions are equivalent, and (ii) contributions of equal shares of income are equivalent. However, system (i) will not in general support a Kantian maxim because individuals with different preferences and/or income[5] will generally want to erect different levels of contribution into a universal law, thus violating the moral equivalence condition. Similarly, system (ii) will not support a Kantian maxim in general if individuals have different income unless they have identical homothetic preferences. If preferences are different, or if incomes are different and the preferences are identical but not homothetic, individuals would want different shares of income to be spent on the public good. Again, this would violate the moral equivalence condition. Yet in the special cases where a Kantian maxim can be defined with respect to either systems of universalization, it prescribes a Pareto efficient level of contribution. The question thus arises: In a general setting where individuals have arbitrarily different preferences and income, will a Kantian maxim always prescribe a Pareto-efficient level of contributions? To address this question we will further restrict the system of universalization to be differentiable.[6]

Definition 6 *A tight system of universalization is* ***differentiable*** *if the* n *one-to-one functions* $\Psi_i^h : S_h \to S_i$ *representing it are differentiable.*

---

[5] Strictly speaking, system (i) is not a well-defined system of universalization if individuals have different incomes because contributions in excess of $\min_i\{\omega_i\}$ do not have a moral equivalent in every strategy sets. We could complete it easily by assuming that equal contributions smaller than $\min_i\{\omega_i\}$ are equivalent but that all contributions greater or equal to $\min_i\{\omega_i\}$ are morally equivalent to each other. However, the resulting system would not be tight.

[6] The *differentiability* assumptions that are imposed both on preferences and on the system of universalization are not essential to the analysis. but they simplify it a great deal. The analysis does, however, depend to some extent upon the *continuity* of the system of universalization which is satisfied under differentiability.

The following lemma is proved in the appendix:

Lemma 5 *Let $(N, \times_{i\in N}[0,\omega_i], < V_i(.) >_{i\in N})$ be a GVPPG in $\mathcal{P}$ and let M be a tight and differentiable system of universalization on $\bigcup_{i\in N}\{i\}\times[0,\omega_i]$. Then a maxim $\mu \subseteq \times_{i\in N}[0,\omega_i]$ satisfies moral equivalence and universalized rationality with respect to M only if the n one-to-one functions $\Psi_i^h : [0,\omega_h] \rightarrow [0,\omega_i]$ (for $i \in N$) representing M are monotonically increasing.*

If any tight and continuous system of universalization is to support a Kantian maxim, it must be representable by monotonically increasing functions. This just means that people see greater contributions by someone as morally equivalent to greater contributions by someone else. This makes intuitive sense since by assumption the public good is not a 'bad' for anyone. Given this, we establish that there is no conflict between Kantian behavior and Pareto-efficiency. More specifically, if a Kantian maxim can be defined with respect to a tight and differentiable system of universalization, then this maxim must prescribe a Pareto-efficient level of contributions. Laffont (1975)'s central result is therefore robust to relaxing the assumption of identical preferences and wealth.

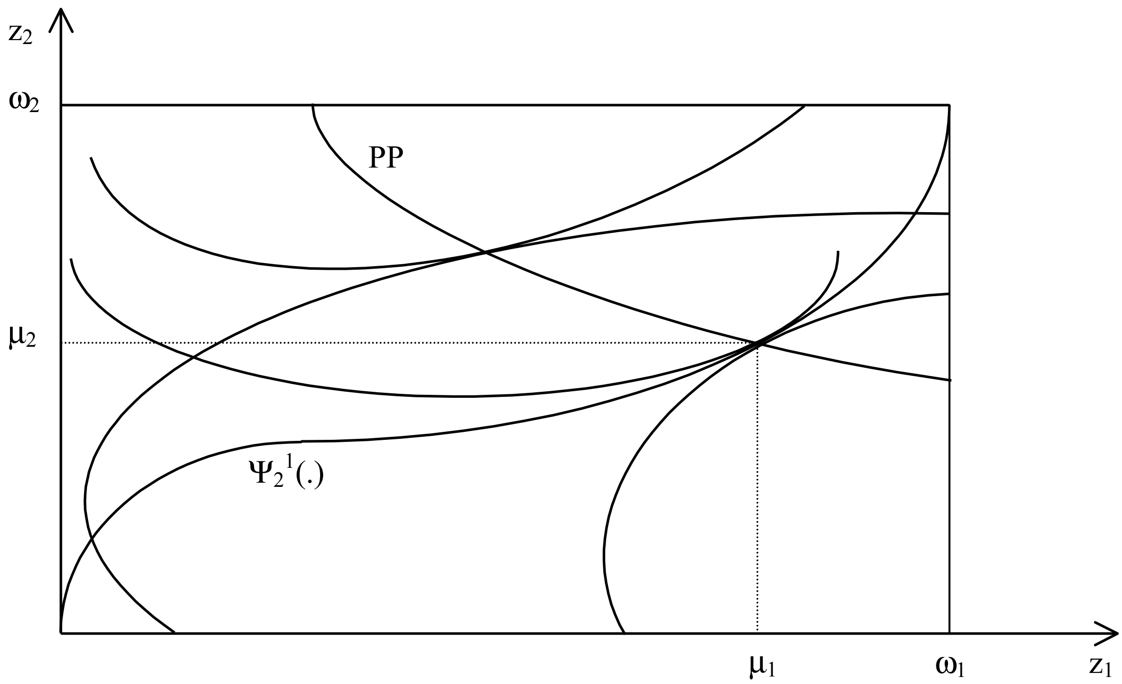
Proposition 1 *Let $(N, \times_{i\in N}[0,\omega_i], < V_i(.) >_{i\in N})$ be a GVPPG in $\mathcal{P}$ and let M be a tight and differentiable system of universalization on $\bigcup_{i\in N}\{i\}\times[0,\omega_i]$. Then if a maxim $\mu \subseteq \times_{i\in N}[0,\omega_i]$ satisfies moral equivalence and universalized rationality with respect to M, it is Pareto-efficient.*

Proposition 1 is illustrated in figure 1 in an economy with two individuals.

<Insert Figure 1 here>

In the (z_1, z_2) plane, the indifference curves of individual 1 are U-shaped, and those of individual 2 are C-shaped. Line PP is the locus of points where the indifference curves of the

Figure 1



two individuals are tangent, and shows the set of Pareto-efficient combinations of contributions. The system of universalization is represented by the (differentiable and monotonically increasing) function $\Psi_2^1(.)$ joining $(0,0)$ to (ω_1, ω_2).

To satisfy universalized rationality the maxim must require each individual to choose a level of contribution where his indifference curve is tangent to this function (possible "corner solutions" are addressed in the proof of the proposition). To satisfy moral equivalence, the contributions of the two individuals must be at the same point on the $\Psi_2^1(.)$ curve. This can only happen if the two indifference curves are tangent to each other at this point, which means that a Kantian maxim must prescribe contributions at the point where $\Psi_2^1(z_1)$ intersects PP.

Proposition 1 is proved in the appendix for the general case of n individuals. It is worth noting to this respect that the geometric illustration of the proof in figure 1 is valid only when there are two individuals. Figure 1 may in effect give the impression to the casual reader that interior Pareto-efficient combinations of contributions are geometrically characterized by the tangency of indifference curves defined in the n-dimensional Euclidean space of *contribution combinations* (i.e. in the $(z_1, z_2, .., z_n)$ space). Yet this geometric intuition is correct only when $n = 2$. As is well-known, for the class of public good provision problems considered here, an interior efficient allocation of a private good and a public good is characterized by the Samuelson condition that the slopes of the indifference curves defined in the $2-$dimensional Euclidian space of *private and public good bundles* (i.e. in the (Z, x_i) space for every individual i) sum to one. If there are only two individuals, Samuelson's condition turns out to be equivalent to a tangency between indifference curves defined in (z_1, z_2) space. But this equivalence does *not* generalize to more than 2 individuals.

3.3 Existence

We now turn to the question of existence of a differentiable and tight system of universalization that could support a Kantian maxim. We refer again to figure 1 for a geometric intuition into the question, keeping in mind the somewhat misleading representation provided by this picture for the case where there are more than 2 individuals. We see that for the system of moral equivalence to support a Kantian maxim, the slope of the function $\Psi_2^1(.)$ representing it must be the same at the point $(\widehat{z}_1, \widehat{z}_2)$ where it intersects PP as that of the indifference curves going through that point. If it is not, then that system of moral equivalence does not support a Kantian maxim. So the problem of finding a tight and differentiable system of universalization that supports a Kantian maxim amounts geometrically, to finding a continuous function connecting $(0,0)$ to (ω_1, ω_2) that separates the indifference curves that are tangent at $(\widehat{z}_1, \widehat{z}_2)$. This problem is quite similar to that of finding a set of personalized Lindahl prices in an economy with a public good. In figure 1, a set of Lindahl prices for individuals 1 and 2 could be represented by a ray starting at the origin that separates their indifference curves somewhere along locus PP. We know from standard existence theorems for Lindahl equilibria in economies with public goods (see e.g. Milleron (1972)) that there would always exists at least one such ray. The difference with the usual problem of separating indifference surfaces by hyperplanes is that, here, the supporting system of universalization must connect continuously $(0,0)$ to (ω_1, ω_2). This imposes an additional constraint on the separation argument which needs to be properly accounted for in the proof.

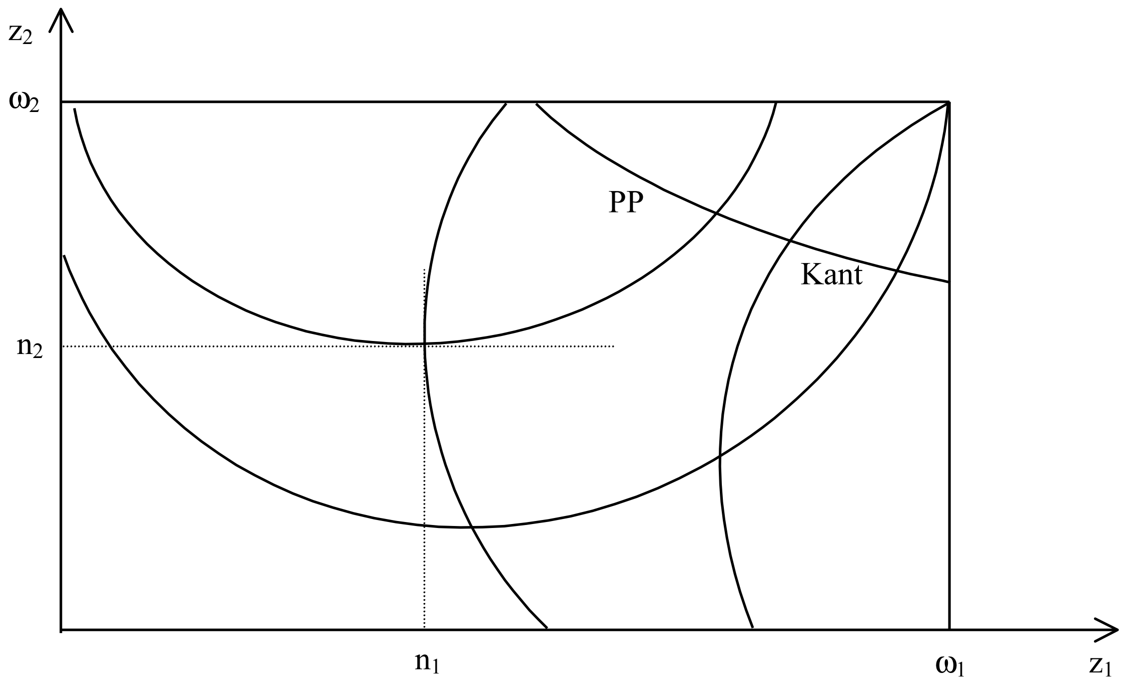
Proposition 2 *Let* $(N, \times_{i\in N}[0, w_i], < V_i(.) >_{i\in N})$ *be a GVPPG in* $\mathcal{P}$ *. Then there exists a tight and continuous system of universalization* M *that supports the Kantian maxim "Contribute your Lindahl equilibrium amount."*

The proof (in the appendix) establishes that a system of universalization, which sees contributions that are in the same ratio as they would be in a Lindahl equilibrium as being morally equivalent over a suitably chosen range, could support a Kantian maxim. It also shows that the system of universalization can be chosen to be essentially piece-wise linear. Note that the system of universalization constructed in the proof is not necessarily the only system of universalization that could support a Kantian maxim. However, since any Kantian maxim is Pareto-efficient (and therefore leads to an allocation of private and public goods that could be supported by Lindahl prices), the local coincidence of the system of universalization with the Lindahl prices must be an element of any proof showing the existence of a tight system of universalization supporting a Kantian maxim.

Proposition 2 is good news for those who believe in the potential usefulness of calling upon individual morality to eliminate free-riding: A Kantian maxim can be defined even when individuals are not identical, provided that the prevailing notion of moral equivalence takes into account individuals' differing preferences and wealth. However, the conditions for this to happen are fairly stringent. In particular, individuals must have enough information to at least be able to figure out what their "Lindahl equilibrium contribution" would be.

While these results suggest that the adoption of a Kantian-like code of conduct by voluntary contributors is normatively desirable (at least on efficiency grounds) and logically possible, they do not say anything about whether the individuals have incentives to obey a Kantian maxim rather than to behave in a non-cooperative fashion (say according to the unique Nash equilibrium of this game). Clearly, by definition of the Nash equilibrium, no individual can unilaterally gain from obeying alone the prescription of a Kantian maxim. But

Figure 2



at least, if *all* individuals could gain by *collectively* obeying such a maxim, then there might be some hope that they could be convinced (perhaps through education or propaganda) to do so. Therefore, we would like to know whether there will always exist a tight system of universalization that supports a Kantian maxim whose outcome Pareto-dominates the Nash equilibrium. Unfortunately, the answer is no. It is easy to construct examples in which *no* Kantian maxim Pareto-dominates the unique Nash equilibrium outcome of the game. Figure 2 gives such an example. Any allocation that could be achieved by universal obedience to a Kantian maxim (the set of such allocations is labeled "Kant" on the picture) is considered worse by individual 1 than what she gets at the Nash equilibrium (n_1, n_2).[7]

<Insert figure 2 about here>

4 Conclusion

The purpose of this paper was to assess the consequences of Kantian morality in a world populated by nonidentical individuals. By Kantian morality we meant throughout the common interpretation in economics of Kant's categorical imperative: undertaking any action that one would want everyone else to undertake. To formalize this, we have defined the concept of a maxim as a joint restriction on the strategy sets of all players in a game. Kantian individuals obeying a maxim are restricted to playing strategies in this set. We have proposed two properties that a maxim must satisfy to qualify as Kantian: it must treat everyone similarly and it must prescribe a combination of actions that is best for everyone under this constraint of similar treatment of everyone. We have defined these two properties

[7] This state of affairs is obviously related to the fact that a Lindahl equilibrium does not always Pareto-dominate the Nash equilibrium of a GVPPG. For more in this question, see Shitovitz and Spiegel (1998).

with respect to the general notion of a system of universalization which defines whether two strategies are morally equivalent or not.

We found that in general, the mere existence of a Kantian maxim cannot be taken for granted when individuals are nonidentical. Some games do not admit a Kantian maxim for all but the trivial system of universalization that considers all strategies to be morally equivalent to each other. In games of voluntary contributions to a public good, the verdict is more positive. We have shown that there will always exist at least one tight and differentiable system of universalization with respect to which a Kantian maxim can be defined. However the conditions for this to happen are fairly stringent. In particular, the system of universalization must take individuals' preferences and wealth into account in a very particular (though not implausible) way.

The other issue we looked at was whether universal obedience to a Kantian maxim, if one exists, would yield a Pareto-efficient outcome. While we noticed that for general games Kantian maxims do not necessarily always prescribe Pareto-efficient strategy combinations, the situation is brighter in games of voluntary provision of a public good. We have shown that a Kantian maxim is always Pareto-efficient. Hence Laffont's original insight that Kantian morality could help overcome the free-rider problem remains valid in a world populated by nonidentical individuals.

5 Appendix

Proof of lemma 1. : Let M be a system of universalization on $\bigcup_{i \in N} \{i\} \times S_i$. For any individual h, define, for $i \in N$, the correspondence $\Psi_i^h : S_h \to S_i$ by $\Psi_i^h(s'') = E_M(h, s'') \cap \{i\} \times S_i$ for every $s'' \in S_h$. To see that Ψ_i^h is onto, we simply note that for every $s \in S_i$, we have $E_M(i, s) \cap \{h\} \times S_h \neq \emptyset$ so that there exists some $s'' \in S_h$ such that $(h, s'')\ M\ (i, s)$ and, by symmetry of M, such that $(i, s) \in E_M(h, s'') \cap \{i\} \times S_i = \Psi_i^h(s'')$. To see that $(i, s)\ M\ (j, s') \Leftrightarrow (i, s) \in \Psi_i^h(s'')$ for every i, $j \in N$, $s \in S_i$, $s' \in S_j$, and $s'' \in \Psi_h^j(s')$ suppose first that $(i, s)\ M\ (j, s')$. Since Ψ_j^h is onto, there is some $s'' \in S_h$ such that $s'' \in \Psi_h^j(s')$. By definition of Ψ_j^h and symmetry of M, we have $(j, s')\ M\ (h, s'')$ and, by transitivity of M, $(i, s)\ M\ (h, s'')$ for all $s'' \in \Psi_h^j(s_j)$. Hence $s \in \Psi_i^h(\Psi_h^j(s''))$. Conversely, let $s \in \Psi_i^h(s'')$ for $s'' \in \Psi_h^j(s')$. By definition of Ψ_h^j, we then have $(h, s'')\ M\ (j, s')$. Since $(i, s) \in E_M(h, s'') \cap \{i\} \times S_i$, we have that $(i, s)\ M\ (h, s'')$ and, by transitivity, $(i, s)\ M\ (j, s')$. ■

Proof of lemma 5. : We note first that, as a continuous and one-to-one function from $[0, \omega_h]$ to $[0, \omega_i]$ (for $i \in N$), Ψ_i^h must be monotonic. It can be monotonically decreasing or monotonically increasing. We wish to show that it must be monotonically increasing in order to support a Kantian maxim $\mu \subseteq \prod_{i \in N} [0, \omega_i]$. Assume therefore that the strict maxim $\mu = (\mu_1, ..., \mu_n)$ satisfies universalized rationality and moral equivalence with respect to the tight system of universalization represented by the n differentiable one-to-one functions $\Psi_i^h : [0, \omega_h] \to [0, \omega_i]$ (for $i \in N$). Since μ satisfies moral equivalence, we have $\mu_i = \Psi_i^h(\mu_h)$ for all i. Moreover for the functions Ψ_i^h (for $i \in N$) to represent a tight system of universalization, we must have, for every $z_h \in [0, \omega_h]$, $\Psi_h^h(z_h) = z_h$ since (by reflexivity) $(h, z_h)\ M\ (h, z_h)$. For

μ to satisfy universalized rationality, we must have, for every individual $i \in N$

$$\mu_h \in \arg \max_{z_h \in [0,\omega_h]} U_i(\sum_{j \in N} \Psi_j^h(z_h), \omega_i - \Psi_i^h(z_h))$$

Given the differentiability assumption and the Kuhn-Tucker theorem, if $\mu_h \in [0, \omega_h[$, it must satisfy the first order condition

$$MRS^i(\sum_{j \in N} \Psi_j^h(\mu_h), \omega_i - \Psi_i^h(\mu_h)) \leq \frac{\frac{\partial \Psi_i^h(\mu_h)}{\partial z_h}}{\sum_{j \in N} \frac{\partial \Psi_j^h(\mu_h)}{\partial z_h}} \tag{1}$$

for all i. Since individual preferences are increasingly strictly monotonic,

$$MRS^i(\sum_{j \in N} \Psi_j^h(\mu_h), \omega_i - \Psi_i^h(\mu_h)) > 0$$

so that (1) can only be satisfied if one of the following two cases holds: (i) for all i $\frac{\partial \Psi_i^h(\mu_h)}{\partial z_h} > 0$ and $\sum_{j \in N} \frac{\partial \Psi_j^h(\mu_h)}{\partial z_h} > 0$, or (ii) For all i $\frac{\partial \Psi_i^h(\mu_h)}{\partial z_h} < 0$ and $\sum_{j \in N} \frac{\partial \Psi_j^h(\mu_h)}{\partial z_h} < 0$. Yet, since $\frac{\partial \Psi_h^h(\mu_h)}{\partial z_h} = 1$, case (ii) must be ruled out so that we must have $\frac{\partial \Psi_i^h(\mu_h)}{\partial z_h} > 0$ and, since the functions Ψ_i^h are monotonic, there must be increasingly so. Suppose now that $\mu_h = \omega_h$ and, by Kuhn-Tucker theorem, that the inequality

$$MRS^i(\sum_{j \in N} \Psi_j^h(\omega_h), \omega_i - \Psi_i^h(\omega_h)) \geq \frac{\frac{\partial \Psi_i^h(\omega_h)}{\partial z_h}}{\sum_{j \in N} \frac{\partial \Psi_j^h(\omega_h)}{\partial z_h}} \tag{2}$$

is satisfied for all i. By contradiction, suppose that for some individual k, the function $\Psi_k^h(.)$ is decreasingly monotonic. Let $N_D = \{i \in N : \Psi_i^h(.)$ is decreasing$\}$ and $N_I = \{i \in N : \Psi_i^h(.)$ is increasing$\}$. We note that the two sets are non empty (the first one by assumption, the second because it contains h). Since by assumption, we have $U_i(\sum_{j \in N_D} \omega_j, \omega_i) > U_i(\sum_{j \in N_I} \omega_j, 0) = U_i(\sum_{j \in N} \mu_j, \omega_i - \mu_i)$ for all $i \in N_I$, we derive immediately the required contradiction of the assumption that μ satisfies universalized rationality. ■

Proof of proposition 1. : Let $\mu \in \times_{i\in N}[0,\omega_i]$ be a (strict) maxim satisfying moral equivalence and universalized rationality with respect to a tight and differentiable (and therefore continuous) system of universalization M. By lemma 5, M must be represented by n monotonically increasing one-to-one differentiable functions $\Psi_i^h : S_h \to S_i$ (for $i \in N$). For μ to satisfy moral equivalence, we must have $\mu_i = \Psi_i^h(\mu_h)$ for all i. For μ to satisfy universalized rationality, we must have, for every individual $i \in N$,

$$\mu_h \in \arg\max_{z_h\in[0,\omega_h]} U_i(\sum_{j\in N}\Psi_j^h(z_h), \omega_i - \Psi_i^h(z_h))$$

We first consider the case where $\mu_h \in]0,\omega_h[$. In this case, μ_h satisfies the first order condition

$$MRS^i(\sum_{j\in N}\Psi_j^h(\mu_h), \omega_i - \Psi_i^h(\mu_h)) = \frac{\frac{\partial\Psi_i^h(\mu_h)}{\partial z_h}}{\sum\limits_{j\in N}\frac{\partial\Psi_j^h(\mu_h)}{\partial z_h}} \tag{3}$$

with (since the functions Ψ_i^h are increasing) $\frac{\partial\Psi_i^h(\mu_h)}{\partial z_h} > 0$ for all i and $\sum\limits_{j\in N}\frac{\partial\Psi_j^h(\mu_h)}{\partial z_h} > 0$. Summing these conditions over the n individuals yields

$$\sum_{i\in N} MRS^i(\sum_{j\in N}\Psi_j^h(\mu_h), \omega_i - \Psi_i^h(\mu_h)) = \sum_{i\in N}\frac{\frac{\partial\Psi_i^h(\mu_h)}{\partial z_h}}{\sum\limits_{j\in N}\frac{\partial\Psi_j^h(\mu_h)}{\partial z_h}} = 1$$

which is the well-known Samuelson condition which, by lemma 4, is sufficient for Pareto efficiency. Suppose now that $\mu_h = 0$ but, by contradiction, suppose that 0^N is not a Pareto efficient vector of contributions. That is, suppose that there exists a vector of contributions $(\widehat{z}_1, ..., \widehat{z}_n) \in \times_{i\in N}[0,\omega_i]$ such that, for all $i \in N$, $U_i(\widehat{Z}, \omega_i - \widehat{z}_i) \geq U_i(0,\omega_i)$ and, for some individual $j \in N$, $U_j(\widehat{Z}, \omega_j - \widehat{z}_i) > U_j(0,\omega_j)$ (where $\widehat{Z} = \sum\limits_{i\in N}\widehat{z}_i$). Let $\widehat{z} \in [0,\omega_h]$ be such that $\sum\limits_{i\in N}\Psi_i^h(\widehat{z}) = \widehat{Z}$. Such a $\widehat{z}$ exists since the function $G : [0,\omega_h] \to [0,\omega]$ defined, for every $z_h \in [0,\omega_h]$, by $G(z_h) = \sum\limits_{i\in N}\Psi_i^h(z_h)$ is one-to-one and continuous. By assumption, for all individual i, we have $U_i(\widehat{Z}, \omega_i - \widehat{z}_i) \geq U_i(0,\omega_i) \geq U_i(\widehat{Z}, \omega_i - \Psi_i^h(\widehat{z}))$ and, for at least one individual j, we have $U_j(\widehat{Z}, \omega_j - \widehat{z}_j,) > U_j(0, \omega_j) \geq U_j(\widehat{Z}, \omega_j - \Psi_j^h(\widehat{z}))$. Since preferences are strictly increasingly monotonic, it follows that $\widehat{z}_i \leq \Psi_i^h(\widehat{z})$ for all i and $\widehat{z}_j < \Psi_j^h(\widehat{z})$ for at least one individual j. Summing these inequalities over all n individuals leads us to the required contradiction that $\widehat{Z} = \sum_{i \in N} \widehat{z}_i < \sum_{i \in N} \Psi_i^h(\widehat{z})$. The case where $\mu_h = \omega_h$ (and therefore, since the $\Psi_i^h(.)$ are monotonic, where $\mu_i = \omega_i$ for all i) can be ruled out at once by condition 1. ■

The following lemma will be used in the existence proof below:

Lemma 6 *Let $(N, \times_{i \in N}[0, w_i], < V_i(.) >_{i \in N}) \in \mathcal{P}$. Then, there exists a tight and differentiable system of universalization M on $\bigcup_{i \in N} \{i\} \times [0, \omega_i]$ represented by a list of n differentiable, monotonically increasing and one-to-one functions $\Psi_i^h : [0, \omega_h]$ mentioned in lemma 2 if and only if there exists n differentiable and monotonically increasing one-to-one functions $G_i : [0, \omega] \to [0, \omega_i]$.*

Proof. For some individual $h \in N$, let $\Psi_i^h : [0, \omega_h] \to [0, \omega_i]$ (for $i \in N$) be the n differentiable, monotonically increasing and one-to-one functions representing the tight and differentiable system of universalization M on $\bigcup_{i \in N} \{i\} \times [0, \omega_i]$. Define $G : [0, \omega_h] \to [0, \omega]$ by $G(z_h) = \sum_{i \in N} \Psi_i^h(z_h)$. It is immediate to verify that G is differentiable, monotonically increasing and one-to-one on its domain and range if the Ψ_i^h are on theirs. Define therefore its inverse $G^{-1} : [0, \omega] \to [0, \omega_h]$ in the usual fashion by $G^{-1}(z) = z_h \Longleftrightarrow z = G(z_h)$ and define the n functions $G_i : [0, \omega] \to [0, \omega_i]$ by $G_i(z) = \Psi_i^h(G^{-1}(z))$. We leave to the reader the task of verifying that the n functions G_i so defined are differentiable, increasingly monotonic and one-to-one. Suppose now that we have n differentiable and monotonically increasing one-to-one functions $G_i : [0, \omega] \to [0, \omega_i]$. Define the system of universalization M on $\bigcup_{i \in N} \{i\} \times [0, \omega_i]$ by (i, z_i) M (j, z_j) (for $i, j \in N$, $z_i \in [0, \omega_i]$ and $z_j \in [0, \omega_j]$) $\Leftrightarrow z_i = G_i(G_j^{-1}(z_j))$ with G_j^{-1} denoting the (one-to-one, differentiable and monotonically increasing) inverse of G_j. We need to prove that M is indeed a system of universalization. Reflexivity of M is trivially established. For symmetry, $(i, z_i)\ M\ (j, z_j) \Leftrightarrow z_i = G_i(G_j^{-1}(z_j)) \iff z = G_i^{-1}(z_i)$ (for $z = G_j^{-1}(z_j)) \iff z_j = G_j(G_i^{-1}(z_i)) \iff (j, z_j)\ M\ (i, z_i)$. For transitivity, assume that $(i, z_i)\ M\ (j, z_j)$ and $(j, z_j)\ M\ (k, z_k)$ for $i, j, k \in N$, $z_i \in [0, \omega_i]$, $z_j \in [0, \omega_j]$ and $z_k \in [0, \omega_k]$. Hence, one has 1) $z_i = G_i(G_j^{-1}(z_j))$ and 2) $z_j = G_j(G_k^{-1}(z_k))$. Let $z = G_k^{-1}(z_k)$. From 2) $z = G_j^{-1}(z_j) = G_k^{-1}(z_k)$. It follows from 1) that $z_i = G_i(G_k^{-1}(z_k))$ and, therefore, that (i, z_i) $M\ (k, z_k)$ as required by transitivity. The verification that, for every individuals i, $j \in N$, and for $z_i \in [0, \omega_i]$, $\#E_M(i, z_i) \cap [0, \omega_j] = 1$ is left to reader. ■

Proof or proposition 2. Let $(z_1^*, ..., z_n^*)$, $(p_1^*, ..., p_n^*)$ and $(u_1^*, ..., u_n^*)$ denote the vectors of contributions, personalized prices and utilities (respectively) associated to a Lindahl equilibrium for this game and let $Z^* = \sum_{i \in N} z_i^*$. Recall that by definition of a Lindahl equilibrium

$$(Z^*, \omega_i - p_i^* Z^*) \in \arg\max_{(Z,x)} U_i(Z, x) \text{ subject to } p_i^* Z + x \leq \omega_i$$

The assumptions imposed on $\mathcal{P}$ guarantee the existence of a Lindahl equilibrium (see e.g., Milleron (1972)). For every individual $i \in N$, let $f_i : \mathbb{R}_+ \to \mathbb{R}$ be the (implicit) function defined by $U_i(Z, f_i(Z)) \equiv u_i^*$. The assumptions imposed on $\mathcal{P}$ also secure the existence of this function. Let $s_i = \frac{\omega_i}{\omega}$ denote the share of the community's wealth owned by individual i. Partition N into sets A, B and C defined by

$$\begin{aligned} A &= \{i \in N : s_i < p_i^*\} \\ B &= \{i \in N : s_i > p_i^*\} \\ C &= \{i \in N : s_i = p_i^*\} \end{aligned}$$

Since $\sum_{i \in N} s_i = \sum_{i \in N} p_i^* = 1$, then $A = \emptyset \Longleftrightarrow B = \emptyset$. Let $\overline{Z}_i$ be defined by

$$\overline{Z}_i = \frac{\omega_i - f_i(\omega)}{p_i^*}$$

Two things can be noted about these $\overline{Z}_i$. First, by condition 1, $U_i(0, \omega_i) > U_i(\omega, 0)$. Moreover, by definition of a Lindahl equilibrium, $U_i(Z^*, \omega_i - z_i^*) \geq U_i(0, \omega_i)$ which, given that $U_i(.)$ is monotonically increasing, implies that $f_i(\omega) > 0$ and, therefore, that $\overline{Z}_i < \frac{\omega_i}{p_i^*}$ for all $i \in N$. Second, one has $\overline{Z}_i > Z^*$. For assuming $\overline{Z}_i \leq Z^*$ would imply assuming $\omega_i - p_i^* Z^* \geq \omega_i - p_i^* \overline{Z}_i = f_i(\omega)$ which is incompatible with the monotonicity of $U_i(.)$. Hence $\overline{Z}_i \in]Z^*, \frac{\omega_i}{p_i^*}[$. Now, let $\overline{a} \leq \#A$ be the number of distinct $\overline{Z}_i$ that one can find in A (that is $\overline{a} = \#\{Z \in [Z^*, \omega] : \exists\ i \in A$ such that $Z = \overline{Z}_i\}$). Write A as $A = A_1 \cup A_2 \cup ... A_{\overline{a}}$ where, for every individuals i, $j \in A$, ($i, j \in A_k$ for some $k \in \{1, ..., \overline{a}\}) \Longleftrightarrow \overline{Z}_i = \overline{Z}_j$ and ($i \in A_k$, $j \in A_{k'}$ for some $k, k' \in \{1, ..., \overline{a}\}$ such that $k' > k) \Longleftrightarrow \overline{Z}_i < \overline{Z}_j$. Label these $\overline{a}$ distinct values of $\overline{Z}_i$ as $\overline{Z}_1, ..., \overline{Z}_{\overline{a}}$,and write $[0, \omega] = \bigcup_{k=0}^{\overline{a}} I_k$ with $I_0 = [0, \overline{Z}_1]$, $I_k = [\overline{Z}_k, \overline{Z}_{k+1}]$ for $k = 1, ..., \overline{a} - 1$, and $I_{\overline{a}} = [\overline{Z}_{\overline{a}}, \omega]$. Using the lemma above, define the n functions $G_i : [0, \omega] \to [0, \omega_i]$ by $G_i(Z) = G_i^k(Z)$ for $Z \in I_k$, $k \in \{0, 1, ..., \overline{a}\}$ with the functions $G_i^k : I_k \to [0, \omega_i]$ defined as follows.

$$G_i^0(Z) = p_i^* Z \text{ for all } i \in N \tag{A1}$$

$$G_i^k(Z) = G_i^{k-1}(Z) \text{ for } i \in N \backslash (A_k \cup \{j(k)\}, \tag{A2}$$

$$G_i^k(Z) = \omega_i - [\frac{\omega_i - p_i^* \overline{Z}_k}{\omega - \overline{Z}_k}][\omega - Z] \text{ for } i \in A_k \tag{A3}$$

$$G_{j(k)}^k(Z) = G_{j(k)}^{k-1}(\overline{Z}_k) + [\frac{\partial G_{j(k)}^{k-1}(\overline{Z}_k)}{\partial Z} + \sum_{i \in A_k} (p_i^* - (\frac{\omega_i - p_i^* \overline{Z}_k}{\omega - \overline{Z}_k})](Z - \overline{Z}_k) \tag{A4}$$

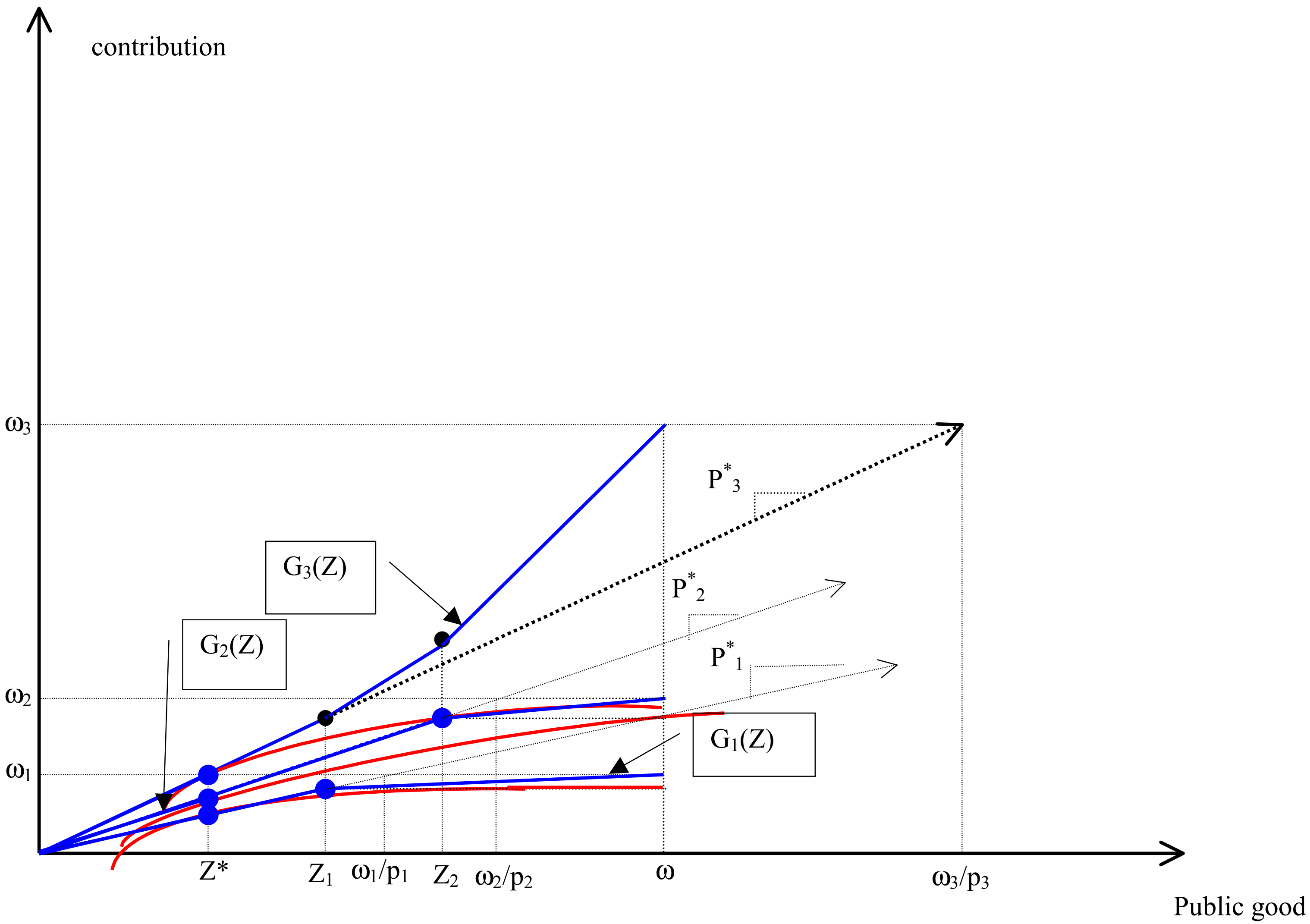

contribution
ω3
ω2
ω1
G3(Z)
G2(Z)
G1(Z)
P*3
P*2
P*1
Z*
Z1
ω1/p1
Z2
ω2/p2
ω
ω3/p3
Public good

Figure 3

if $k = 1, ..., \overline{a} - 1$ for $j(k) \in B$ chosen in such a way that $\omega_{j(k)} - G_{j(k)}^{k-1}(\overline{Z}_k) \geq \omega_j - G_j^{k-1}(\overline{Z}_k)$ for all $j \in B$ and

$$G_i^{\overline{a}}(Z) = \omega_i - [\frac{\omega_i - G_i^{\overline{a}-1}(\overline{Z}_{\overline{a}})}{\omega - \overline{Z}_{\overline{a}}}][\omega - Z] \text{ for all } i \in N \tag{A5}$$

Figure 3 illustrates how the functions $G_i(.)$ are constructed in a case where there are 3 individuals and 2 distinct values of $\overline{Z}_i$ (labeled Z_1 and Z_2 on the picture). Note carefully that figure 3 takes place in the *public good and contribution level space* so that the utility levels of the three individuals are increasing in the south east direction.

<**Insert figure 3 about here**>

We note immediately that $G_i(0) = 0$, $G_i(\omega) = G_i^{\overline{a}}(\omega) = \omega_i$. The fact that $G_i^k(.)$ is continuous and strictly increasing on I_k (for $k = 0, ..., \overline{a}$) is also clear except, perhaps, for the functions $G_{j(k)}^k(.)$. In order to prove that these functions are increasing monotonically with respect to Z on their domain, we need to show that $\frac{\partial G_{j(k)}^k(.)}{\partial Z} = [\frac{\partial G^{k-1}(\overline{Z}_k)}{\partial Z} + \sum_{i \in A_k} (p_i^* - (\frac{\omega_i - p_i^* \overline{Z}_k}{\omega - \overline{Z}_k})] > 0$. For this, we simply note that, for all $i \in A$,

$$p_i^* - (\frac{\omega_i - p_i^* Z}{\omega - Z}) = \frac{p_i^* \omega - \omega_i}{\omega - Z} > 0$$

for all $Z \in [0, \frac{\omega_i}{p_i^*}]$ and, therefore, that, by the very definition of $\frac{\partial G_{j(k)}^{k-1}(\overline{Z}_k)}{\partial Z}$:

$$0 < p_{j(k)}^* \leq \frac{\partial G_{j(k)}^{k-1}(\overline{Z}_k)}{\partial Z} \leq p_{j(k)}^* + \sum_{k' \in \{1,...,k-1\}} \sum_{i \in A_{k'}} (p_i^* - (\frac{\omega_i - p_i^* \overline{Z}_k}{\omega - \overline{Z}_k})]$$

We now prove that $G_i^k(\overline{Z}_k) = G_i^{k-1}(\overline{Z}_k)$ for all $k = 1, ..., \overline{a}$. This is clear by definition of $G_i^k(.)$, at every $k = 1, ..., \overline{a} - 1$, for $i \in N \backslash (A_k \cup \{j(k)\}$ and for $G_{j(k)}^k(.)$. For $i \in A_k$, we note that $G_i^{k'}(Z) = G^0(Z) = p_i^* Z$ for all $Z \in \bigcup_{k'<k} I_{k'}$, for all $k' < k$. Hence $G_i^k(\overline{Z}_k) = \omega_i - [\frac{\omega_i - p_i^* \overline{Z}_k}{\omega - \overline{Z}_k}][\omega - \overline{Z}_k] = p_i^* \overline{Z}_k = G_i^{k-1}(\overline{Z}_k)$ for $i \in A_k$. The fact that $G_i^{\overline{a}}(Z_{\overline{a}}) = G_i^{\overline{a}-1}(\overline{Z}_{\overline{a}})$

for all $i \in N$ is immediate. $G_i(.) : [0, \omega] \rightarrow [0, \omega_i]$ being a continuous and increasing function satisfying $G_i(0) = 0$ and $G_i(\omega) = \omega_i$, it is therefore one-to-one.

We now show that $\sum_{i \in N} G_i(Z) = Z$ for all Z. The equality is trivially true for $Z \in I_0$. By finite induction, let us show that if $\sum_{i \in N} G_i(Z) = Z$ holds for $Z \in I_k$, it also holds for $Z \in I_{k+1}$, for every $k = 0, ..., \overline{a} - 1$. For this, we exploit the fact that, for all $k \in \{0, ..., \overline{a}\}$, the functions $G_i^k(.)$ are affine on I_k and can be written as $G_i^k(Z) = a_i^k + b_i^k Z$ for $Z \in I_k$ for some $a_i^k \in \mathbb{R}$ and $b_i^k \in \mathbb{R}_{++}$. This implies that $G_i^k(Z) = G_i^k(\overline{Z}_k) + b_i^k(Z - \overline{Z}_k)$ for $Z \in I_k$ $(k = 0, ..., \overline{a})$. Since, as shown above, $G_i^k(\overline{Z}_k) = G_i^{k-1}(\overline{Z}_k)$ for $k = 1, ..., \overline{a}$, and since by the inductive hypothesis, $\sum_{i \in N} G_i^{k-1}(\overline{Z}_k) = \overline{Z}_k$, $\sum_{i \in N} G_i^k(Z) = Z$ for $Z \in I_k$ if and only if $\sum_{i \in N} b_i^k = 1$. Consider therefore I_k for $k = 1, ..., \overline{a} - 1$. Clearly

$$\sum_{i \in N} b_i^k = \sum_{i \in N/A_k \cup \{j(k)\}} b_i^k + \sum_{i \in A_k} b_i^k + b_{j(k)}^k \tag{B}$$

Moreover, as indicated by (A1)-(A3), $b_i^{k-1} = p_i^*$ for $i \in A_k$, $b_i^k = b_i^{k-1}$ for $i \in N/(A_k \cup \{j(k)\})$ and $b_{j(k)}^{k-1} = \frac{\partial G_{j(k)}^{k-1}(\overline{Z}_k)}{\partial Z}$. Hence, using these conventions and the definitions of the b_i^k contained in (A1)-(A4), one can write (B) as

$$\sum_{i \in N} b_i^k = \sum_{i \in N/A_k \cup \{j(k)\}} b_i^{k-1} + \sum_{i \in A_k} [\frac{\omega_i - p_i^* \overline{Z}_k}{\omega - \overline{Z}_k}] + [b_{j(k)}^{k-1} + \sum_{i \in A_k} (b_i^{k-1} - (\frac{\omega_i - p_i^* \overline{Z}_k}{\omega - \overline{Z}_k})] = \sum_{i \in N} b_i^{k-1} = 1$$

under the inductive hypothesis. For the interval $I_{\overline{a}}$, we note simply that, using (A5)

$$\sum_{i \in N} b_i^k = \sum_{i \in N} [\frac{\omega_i - G_i^{\overline{a}-1}(\overline{Z}_{\overline{a}})}{\omega - \overline{Z}_{\overline{a}}}] = 1$$

under the induction hypothesis. Now let $\mathcal{M} = \{(z_1, ..., z_n) \in \times_{i \in N}[0, \omega_i] : (i, z_i)\ M\ (j, z_j)$ for all $i, j \in N\}$ be the set of all vectors of morally equivalent levels of contribution for the system of moral universalization induced from the functions $G_i(.)$ as per the above lemma.

Using the definition of M of the above lemma, it is immediate to see that this set $\mathcal{M}$ can be equivalently defined as $\mathcal{M} = \{(z_1, ..., z_n) \in \times_{i \in N}[0, \omega_i] : \exists\ Z \in [0, \omega]$, s. to $z_i = G_i(Z)\ \forall$ $i \in N\}$.

We now show that the system of universalization induced by the functions $G_i(.)$ supports the Kantian maxim: contribute z_i^*. It is immediate that $(z_1^*, ..., z_n^*) \in \mathcal{M}$ and, for this reason, that it satisfies moral equivalence with respect to M. In order to establish that it also satisfies universalized rationality, it suffices to prove that $Z^* \in \arg\max_{Z \in [0,\omega]} U_i(Z, \omega_i - G_i(Z))$ for all $i \in N$. By definition of a Lindahl equilibrium, $Z^* \in \arg\max_{Z \in [0,\overline{Z}_1]} U_i(Z, \omega_i - G_i(Z))$ $= \arg\max_{Z \in [0,\overline{Z}_1]} U_i(Z, \omega_i - p_i^* Z)$ for all i. By construction of the functions $G_i(.)$, an individual $i \in A_k$ who universalizes her behavior has access to bundles $(x, Z) \in \mathbb{R}_+^2$ satisfying $x =$ $\omega_i - p_i^* Z$ for $Z \in [0, \overline{Z}_k]$ and $x = [\frac{\omega_i - p_i^* \overline{Z}_k}{\omega - \overline{Z}_k}][\omega - Z]$ for $Z \in [\overline{Z}_k, \omega]$. Since $[\frac{\omega_i - p_i^* \overline{Z}_k)}{\omega - \overline{Z}_k}][\omega -$ $\overline{Z}_k] = f_i(\omega)$, monotonicity of preferences implies that $U_i(Z, (\frac{\omega_i - p_i^* \overline{Z}_k}{\omega - \overline{Z}_k})(\omega - Z)) < u_i^*$. Hence $Z^* \in \arg\max_{Z \in [0,\omega]} U_i(Z, \omega_i - G_i(Z))$ for $i \in A_k$ for every $k \in \{1, ..., \overline{a}\}$. For $i \in B \cup C$, one can check that $G_i(Z) \geq p_i^* Z$ for all $Z \in [0, \omega]$ and, therefore, that all bundles $(Z, \omega_i - G_i(Z))$ satisfying $Z \in [0, \omega]$ are revealed weakly worse than $(Z^*, \omega_i - G_i(Z^*))$.

Now, the functions $G_i(.)$ defined in this proof are not differentiable at $\overline{Z}_k$ (for $k \in$ $\{1, ..., \overline{a}\}$. It is a simple (if not slightly cumbersome) matter to make them differentiable by connecting the $\overline{a}$ pairs of points $(\overline{Z}_k - \varepsilon, G_i(Z_k - \varepsilon))$ to $(\overline{Z}_k + \varepsilon, G_i(Z_k + \varepsilon))$ (for $\varepsilon > 0$) as small as needed by differentiable polynomials whose derivative would coincide with $G_i(.)$ at every of these two points. These polynomials will only concern, at each $k \in \{1, ..., \overline{a}\}$, the members of A_k and the individual $j(k)$ defined above and, as functions from $(\overline{Z}_k - \varepsilon, \overline{Z}_k + \varepsilon)$ to $(G_i(Z_k - \varepsilon)), G_i(Z_k + \varepsilon))$, will be monotonically increasing and satisfy all the required properties. We omit the details of the construction. ∎

References

Bergstrom, T., "On the Evolution of Altruistic Ethical Rules for Siblings," *American Economic Review*, 1995, *85*, 58–81.

——, **L. Blume, and H. Varian**, "On the Private Provision of Public Goods," *Journal of Public Economics*, 1986, *29*, 25–49.

Bernheim, D., "Rationalizable Strategic Behavior," *Econometrica*, 1984, *52*, 1007–1028.

Binmore, K., *Playing Fair*, Cambridge, MA: MIT Press, 1994.

Bordignon, M., "Was Kant right? Voluntary Provision of Public Goods under the Principle of unconditionnal commitment," *Economic Notes*, 1990, *3*, 342–372.

Collard, D. A., *Altruism and the Economy*, Oxford: Martin Robertson, 1978.

Cornes, R. and R. Sandler, "The Theory of Public Goods: Non-Nash Behavior," *Journal of Public Economics*, 1984, *18*, 367–379.

DeJasey, A., *Social Contract and Free Ride: A Study of the Public Goods Problem*, Oxford: Clarendon, 1989.

Etzioni, A., "Toward a Kantian Socio-economics," *Review of Social Economy*, 1987, *45*, 37–47.

Guttman, J., "Understanding Collective Action: Matching Behavior," *American Economic Review: Papers and Proceedings*, 1978, *68*, 346–358.

Kant, I., *Fundamental Principles of the Metaphysics of Ethics*, 3rd ed., London: Longmans, 1907. English Translation by T. K. Abott.

Laffont, J. J., "Macroeconomic Constraints, Economic Efficiency and Ethics: an Introduction to Kantian Economics," *Economica*, 1975, *42*, 430–437.

Milleron, J.-C., "Theory of Value with Public Goods: A Survey Article," *Journal of Economic Theory*, 1972, *5*, 419–477.

Nida-Rumelin, J., "Practical Reason, Collective Rationality and Contractarianism," in D. Gautier and R. Sugden, eds., *Rationality, Justice and the Social Contract: Themes from Moral by Aggreement*, Harvester/wheatsheaf, 1993.

Roemer, J. E., *Theories of Justice*, Cambridge Massachussets: Harvard University Press, 1996.

Samuelson, P. A., "The Pure Theory of Public Expenditure," *The Review of Economics and Statistics*, 1954, *64*, 387–389.

Sandler, T., *Collective Action: Theory and Application*, Ann Arbor, MI: Michigan University Press, 1992.

—— **and J. Posnett**, "The Private Provision of Public Goods: A Perspective on Neutrality," *Public Finance Quarterly*, 1992, *19*, 22–42.

Shitovitz, Benyamin and Menahem Spiegel, "Cournot-Nash and Lindahl Equilibria in Pure Public Good Economies," *Journal Of Economic Theory*, 1998, *83*, 1–18.

Sugden, R., "Reciprocity: the Supply of Public Goods through Voluntary Contributions," *Economic Journal*, 1984.

Wolfelsperger, A., "Sur l'Existence d'une Solution 'Kantienne' du Problème des Biens Collectifs," *Revue Economique*, 1999, *50*, 879–902.

Young, D. J., "A 'Fair Share' Model of Public Good Provisions," *Journal of Economic Behavior and Organization*, 1989, *11*, 137–147.