



HAL
open science

Samaritan vs Rotten Kid: Another Look

Bouwe R. Dijkstra

► **To cite this version:**

Bouwe R. Dijkstra. Samaritan vs Rotten Kid: Another Look. *Journal of Economic Behavior and Organization*, 2007, 64 (1), pp.91. 10.1016/j.jebo.2006.06.006 . hal-00531861

HAL Id: hal-00531861

<https://hal.science/hal-00531861>

Submitted on 4 Nov 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Accepted Manuscript

Title: Samaritan vs Rotten Kid: Another Look

Author: Bouwe R. Dijkstra

PII: S0167-2681(07)00007-8
DOI: doi:10.1016/j.jebo.2006.06.006
Reference: JEBO 2058



To appear in: *Journal of Economic Behavior & Organization*

Received date: 24-9-2004
Accepted date: 20-6-2006

Please cite this article as: Dijkstra, B.R., Samaritan vs Rotten Kid: Another Look, *Journal of Economic Behavior and Organization* (2007), doi:10.1016/j.jebo.2006.06.006

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Samaritan vs Rotten Kid: Another Look

Bouwe R. Dijkstra*

University of Nottingham

January 2006

Abstract. We set up a two-stage game with sequential moves by one altruist and n selfish agents. The Samaritan's dilemma (rotten kid theorem) states that the altruist can only reach her first best when the selfish agents move after (before) the altruist. We find that in general, the altruist can reach her first best when she moves first if and only if a selfish agent's action marginally affects only his own payoff. The altruist can reach her first best when she moves last if and only if a selfish agent cannot manipulate the price of his own payoff.

JEL Classification: D64

Key words: Altruism, rotten kid theorem, Samaritan's dilemma

Lear: I gave you all—

Regan: And in good time you gave it.

(*King Lear*, Act II, Scene IV)

Correspondence: Bouwe R. Dijkstra, School of Economics, University of Nottingham, Nottingham NG7 2RD, UK. Tel: +44 115 8467205, Fax: +44 115 9514159, email: bouwe.dijkstra@nottingham.ac.uk

*I thank Arye Hillman, Gordon Tullock and Heinrich Ursprung for pointing me toward the subject of this paper. I thank Richard Cornes, Hans Gersbach, Matthew Haag, Johan Lagerlöf, Andreas Lange, Markus Lehmann, Andries Nentjes, Till Requate, Bert Schoonbeek, Perry Shapiro, Claudio Zoli and two anonymous referees for valuable comments. Special thanks are due to Ted Bergstrom for essential clarifications. Any remaining errors are my own.

Samaritan vs Rotten Kid: Another Look

1 Introduction

However much we care about other people, we do not wish to invite them to take advantage of our charity. The economic theory of altruism¹ offers two conflicting pieces of strategic advice: the rotten kid theorem (Becker 1974, 1976) and the Samaritan's dilemma (Buchanan 1975). In a single-round model with sequential moves by an altruistic agent (the Samaritan or the parent) and a selfish agent (the parasite or the kid), the contradiction between the two can be stated as follows.

The rotten kid theorem states that the parent can only reach her first best when she moves after the kid. The intuition is that the kid will only act unselfishly if the parent can reward him afterward. The Samaritan's dilemma, on the other hand, states that the Samaritan can only reach his first best when he moves before the parasite. Here, the intuition is that the parasite cannot manipulate the Samaritan's actions when the Samaritan moves first.

In this paper we identify the restrictions on the agents' payoff functions for either result to hold. For the altruist to reach her first best when she moves first, a selfish agent's actions should on balance affect only his own payoff; then there are no externalities to his actions. For the altruist to reach her first best when she moves last, the selfish agents should not be able to manipulate the price of their payoffs to the altruist (i.e. the altruist's trade-off between her own and the selfish agents' payoffs). Then the selfish agents will maximize total payoff. They will benefit from this themselves, because their payoffs are normal goods to the altruist.

Our result for the Samaritan's dilemma is new. For the rotten kid theorem, Bergstrom (1989) has performed a similar analysis. His model is a special version of our general setup. Whereas we do not restrict the nature of the altruist's actions, Bergstrom assumes she distributes a certain amount of money among the selfish agents. Removing this restriction results in a more general condition for the rotten kid theorem. Bergstrom also claims

¹For the evolutionary roots of altruism, see Henrich (2004) and the comments on this paper in the *Journal of Economic Behavior and Organization* 53 (1) Special Issue on evolution and altruism.

that the payoff condition is necessary only when money is important enough. We shall demonstrate that this additional condition is not needed.

Cornes and Silva (1999) have found another condition for the rotten kid theorem to hold in Bergstrom’s framework. We shall see that this condition applies only in Bergstrom’s framework and that there are no additional solutions.

However peripheral to economics the study of altruism may seem, there is in fact an application that takes us to the very heart of the discipline (Munger 2000). Regarding the welfare-maximizing government as an altruist and the private agents as selfish agents, we have a framework for a policy game. This framework allows us to study how the government can shape incentives such that private actions maximize social welfare. Chari et al. (1989) and Cubitt (1992) have addressed this issue using a similar framework. The present paper offers new insights into this question.

The focus of this paper is on the attainment of the altruist’s first best. Another interesting question is whether a particular sequence of moves leads to a Pareto-efficient outcome. Obviously, the altruist’s first best is a Pareto-efficient outcome. Moreover, it can be shown that when the selfish agents move first, the outcome is Pareto efficient if *and only if* it is the altruist’s first best.²

The rest of this paper is organized as follows. In Section 2, we introduce the Samaritan’s dilemma and the rotten kid theorem in simple two-agent setups where they are known to hold. In Section 3, we set up a single-round game with n selfish agents, deriving the conditions for the Samaritan’s dilemma and the rotten kid theorem to hold. In Section 4, we discuss Bergstrom’s game as well as Bergstrom’s own and Cornes and Silva’s conditions for the rotten kid theorem. We conclude with Section 5.

2 Introductory examples

2.1 Samaritan’s dilemma

The Samaritan’s dilemma is due to Buchanan (1975) who discusses a game between an altruistic Samaritan and a selfish parasite.³ He shows that the Samaritan can reach his

²Details are available from the author upon request.

³Buchanan distinguishes between the active and the passive Samaritan’s dilemma. We only discuss the passive Samaritan’s dilemma here. The passive Samaritan’s preferences are reconcilable with a payoff

first best when he moves before the parasite, but not when he moves after the parasite. In this subsection, we shall present a continuous version of the game.⁴

The Samaritan maximizes his objective function $W(U_0, U_1)$, which is increasing in his own payoff U_0 and the parasite's payoff U_1 : $W_k \equiv \partial W / \partial U_k > 0, k = 0, 1$. The parasite maximizes his own payoff U_1 . The Samaritan's own payoff U_0 depends only on his donation y to the parasite, so that we can simply set $U_0 = -y$. The parasite's payoff depends on his work effort x and on the Samaritan's donation y . The parasite's payoff function $U_1(y, x)$ has the following properties:

- $\partial U_1 / \partial y > 0, \partial^2 U_1 / \partial y^2 \leq 0$. The parasite's marginal payoff of money is positive and decreasing.
- $\partial U_1 / \partial x > [<] 0$ for $x < [>] x^*(y)$. Given the Samaritan's donation y , there is an optimal work effort $x^*(y)$ for the parasite, where the marginal payoff of extra money earned equals the marginal payoff of leisure.
- $\partial^2 U_1 / \partial y \partial x < 0$. An increase in the parasite's effort decreases his marginal payoff of money. This is because the parasite earns money for his work, and his marginal payoff of money is decreasing.

The first order conditions for the Samaritan's first best are, with respect to y and x , respectively

$$W_0 = W_1 \frac{\partial U_1}{\partial y} \quad (1)$$

$$\frac{\partial U_1}{\partial x} = 0. \quad (2)$$

We shall now see that the Samaritan can always reach his first best when he moves first, but he can never reach his first best when he moves last. When the Samaritan moves first, his payoff function that only depends on his donation. The active Samaritan's payoff, on the other hand, must also depend on the parasite's action. This is because when the Samaritan donates, he prefers the parasite to go to work although the parasite prefers to stay in bed. Schmidtchen (1999) analyzes the active Samaritan's dilemma.

⁴Jürges (2000) analyzes this game for specific functional forms of W . Bergstrom (p.1140-1) analyzes a similar game, where a parent distributes money after his "lazy rotten kids" have set their work efforts. Neither Bergstrom nor Jürges identify the game with the Samaritan's dilemma.

first, the parasite sets x in stage two to maximize his own payoff:

$$\frac{\partial U_1}{\partial x} = 0.$$

This condition is identical to the first order condition (2) for the Samaritan's first best with respect to x . Thus, in stage one, the Samaritan can set y according to his first best condition (1): The Samaritan can always reach his first best when he moves first.

The intuition is that the parasite sets the work effort that maximizes his own payoff, taking the Samaritan's donation as given. Since the parasite's work effort affects only his own payoff, the parasite takes the full effect of his decision into account. There is no externality, and the Samaritan's first best is implemented.

When the parasite moves first, the Samaritan sets y according to (1) in stage two. In stage one, the parasite sets the x that maximizes his own payoff, taking into account that his choice of x affects the Samaritan's choice of y in stage two:

$$\frac{dU_1}{dx} \equiv \frac{\partial U_1}{\partial x} + \frac{\partial U_1}{\partial y} \frac{dy}{dx} = 0.$$

This corresponds only to the Samaritan's first order condition (2) for x when $dy/dx = 0$ (i.e. the donation reaches its maximum) in the optimum. In order to find the expression for dy/dx in the optimum, we totally differentiate the Samaritan's first order condition for y (1) with respect to x and substitute (2):

$$\frac{dy}{dx} = \frac{W_1 \left(\frac{\partial^2 U_1}{\partial y \partial x} \right)}{-W_{00} + (W_{10} + W_{01}) \frac{\partial U_1}{\partial y} - W_{11} \left(\frac{\partial U_1}{\partial y} \right)^2 - W_1 \frac{\partial^2 U_1}{\partial y^2}} < 0. \quad (3)$$

The numerator in (3) is negative because $W_1 > 0$ and $\partial^2 U_1 / \partial y \partial x < 0$. The denominator is positive because this is the second order condition $\partial^2 W / \partial y^2 < 0$.

Thus, the parasite gets more money from the Samaritan, the less he works. As a result, the parasite will work less than the Samaritan would like him to. The Samaritan cannot reach his first best when he moves after the parasite. Intuitively, the less money the parasite earns, the needier he is and the more money he will get from the Samaritan. When the parasite moves first, he can extort money from the Samaritan by working less.

2.2 Rotten kid theorem

In order to introduce the rotten kid theorem, we analyze the simple game discussed by Becker (1974, 1976) and commented upon by Hirshleifer (1977). The game is between an altruistic parent and a selfish kid. The kid undertakes an action that affects his own as well as the parent's income. The parent can give money to the kid. We shall see that in general, the parent cannot reach her first best when she moves first, but she can always reach her first best when she moves after the kid.⁵

Denote the kid's action by x and the parent's transfer by y . Since the only commodity involved is income, we can equate the parent's and kid's payoffs, U_0 and U_1 respectively, with income and write them in the additively separable form:

$$U_0 = -y + b_0(x) \quad U_1 = y + b_1(x). \quad (4)$$

Here, $b_k(x)$, $k = 0, 1$, is the effect of the kid's action on the income of the parent and the kid, respectively.

The selfish kid maximizes his own payoff U_1 . The parent maximizes her objective function $W(U_0, U_1)$ with $W_k \equiv \partial W / \partial U_k > 0$, $k = 0, 1$. The first order conditions for the parent's first best are, with respect to y and x ,

$$W_0 = W_1 \quad (5)$$

$$W_0 b'_0 + W_1 b'_1 = 0. \quad (6)$$

Substituting (5) into (6),

$$b'_0 + b'_1 = 0. \quad (7)$$

This implies that in the parent's first best, family income $U_0 + U_1 = b_0 + b_1$ is maximized.

When the parent moves first, the kid will set $b'_1 = 0$. In general, this does not correspond to the parent's first order condition (7). When the kid moves last, he will maximize his own income instead of family income.

Now we let the kid move first. In stage two, the parent will set the transfer y that maximizes W , according to (5). In stage one, the kid sets the x that maximizes his

⁵In fact, Becker (1974, 1976) himself does not discuss the order of moves. Citing Shakespeare's *King Lear*, Hirshleifer was the first to point out that the parent's first best is implemented only when the kid moves first.

income, taking into account that his action affects the parent's transfer:

$$\frac{dU_1}{dx} \equiv \frac{dy}{dx} + b'_1 = 0. \quad (8)$$

The value of dy/dx follows from the total differentiation of the parent's first order condition (5) with respect to x :

$$(W_{00} - W_{10}) \left(-\frac{dy}{dx} + b'_0 \right) = (W_{11} - W_{01}) \left(\frac{dy}{dx} + b'_1 \right). \quad (9)$$

By the kid's first order condition (8), the second term between brackets on the RHS of (9) is zero. Thus, the second term between brackets on the LHS of (9) must be zero:

$$\frac{dy}{dx} = b'_0.$$

Substituting this into the kid's first order condition (8), we see that it is equivalent to the parent's first best condition (7): the kid effectively maximizes family income.

Thus, the parent always reaches her first best when she moves after the kid. Bernheim et al. (1985) were the first to note that this result follows from the assumption that there is only one commodity, namely income. The intuition, due to Bergstrom, is that when there is only one commodity, say income, we can identify payoff with income. The kid cannot manipulate the price of his income in terms of the parent's income, because it is always unity. Then the parent and the kid agree that it is a good thing to maximize aggregate income. It is clear that the parent will want to maximize family income. However, as Becker (1974) already notes, the kid will only want to maximize family income if he benefits from that himself, that is if his payoff is a normal good to the parent.

3 A general analysis

3.1 The model

In this section, we analyze a model with one altruistic agent and n selfish agents. We shall see under which conditions the Samaritan's dilemma and the rotten kid theorem hold.

There are $n + 1$ agents, indexed by $k = 0, \dots, n$. Agent 0 is the altruist, and agents i , $i = 1, \dots, n$, are the selfish agents. Agent i controls the variable x_i . Agent 0 can make

a contribution y_i to each agent i 's payoff U_i . Thus $\partial U_i/\partial y_i > 0$ and $\partial U_i/\partial y_j = 0$ for all $i, j = 1, \dots, n, i \neq j$, by definition.

The vector $\mathbf{y} = (y_1, \dots, y_n)$ must be feasible. The lower bound is $\mathbf{y} = \mathbf{0}$: agent 0 can only give to the other agents; she cannot take away from them. There is also an upper bound to \mathbf{y} , which follows from the restriction that agent 0 has only a limited amount of time, money, or whatever the nature of \mathbf{y} , to give to the others. The exact formulation of the upper bound depends on the nature of \mathbf{y} . We assume that neither the upper nor the lower bound are binding constraints on the equilibria.

When agent 0 ultimately gives money (for instance) to the selfish agents, this does not mean that y_i has to be stated as a certain amount of money. Instead, y_i could take the form of a subsidy on behavior from which other agents benefit (e.g. chores).⁶ Obviously, the distinction between unconditional and incentive payment is only relevant when the altruist moves first.

Agent 0's payoff has the form $U_0(\mathbf{y}, \mathbf{x})$, which is continuous and twice differentiable, with $\mathbf{x} = (x_1, \dots, x_n)$. Agent i 's payoff has the form $U_i(y_i, \mathbf{x})$, which is continuous and twice differentiable with $\partial^2 U_i/\partial x_i^2 \leq 0$. Each agent $i, i = 1, \dots, n$, maximizes his own payoff. Agent 0 maximizes her objective function $W(\mathbf{U})$, which is continuous and twice differentiable with $\mathbf{U} = (U_0, \dots, U_n)$, $W_k \equiv \partial W/\partial U_k > 0, k = 0, \dots, n$.

Let us now determine the first-best outcome for agent 0. We assume that the first best is characterized by a unique interior solution. Differentiating $W(\mathbf{U})$ with respect to y_i , and x_i respectively, $i = 1, \dots, n$, we find

$$W_0 \frac{\partial U_0}{\partial y_i} + W_i \frac{\partial U_i}{\partial y_i} = 0 \quad (10)$$

$$\sum_{k=0}^n W_k \frac{\partial U_k}{\partial x_i} = 0. \quad (11)$$

Whatever agent 0's precise preferences, her first best will always be on the payoff possibility frontier *PPF*.

⁶The altruist can also use incentive payments to deal with asymmetric information (Cremer and Pestieau 1996).

Definition 1 Let $(\mathbf{x}^*, \mathbf{y}^*)$ be a set of feasible actions such that there is no other set $(\mathbf{x}', \mathbf{y}')$ of feasible actions with $U_k(\mathbf{x}', \mathbf{y}') \geq U_k(\mathbf{x}^*, \mathbf{y}^*)$ for all k and $U_k(\mathbf{x}', \mathbf{y}') > U_k(\mathbf{x}^*, \mathbf{y}^*)$ for some k , $k = 0, \dots, n$.

Each $(\mathbf{x}^*, \mathbf{y}^*)$ set implements a payoff vector $\mathbf{U}^* \equiv (U_0(y^*, \mathbf{x}^*), U_1(y_1^*, \mathbf{x}^*), \dots, U_n(y_n^*, \mathbf{x}^*))$. The Payoff Possibility Frontier *PPF* is the set of all \mathbf{U}^* .

In the following, we shall study the effect of sequential moves. The agents i , $i = 1, \dots, n$, always move simultaneously. In subsection 3.2, we see what happens when agent 0 moves before agents i . In subsection 3.3, we analyze the case where the agents i move before agent 0. We assume these games have unique interior solutions. We will derive the conditions for these sequences of moves to result in agent 0's first best for all $W(\mathbf{U})$. The conditions will thus be on the payoff functions \mathbf{U} . We are looking for the necessary and sufficient local restrictions on \mathbf{U} under which the first order conditions of the subgame perfect equilibrium are identical to the first order conditions (10) and (11) of agent 0's first best. The local nature of the restrictions means that they must hold on the Payoff Possibility Frontier, since any altruistic agent's first best must be on the *PPF*. We assume that the second order conditions are satisfied.

In the comprehensive interpretation of the Samaritan's dilemma and the rotten kid theorem, they have not only a positive side to them (agent 0 can reach her first best under one sequence of moves), but also a negative side: Agent 0 cannot reach her first best under the other sequence. The relation between the two versions is straightforward: The comprehensive Samaritan's dilemma (rotten kid theorem) holds if and only if the positive Samaritan's dilemma (rotten kid theorem) holds and the positive rotten kid theorem (Samaritan's dilemma) does not hold.

3.2 Agent 0 moves first

In this subsection, we derive the equilibrium for the game where agent 0 moves before agents i , and we see when this equilibrium corresponds to the first best for agent 0. Thus, we shall derive the condition for the positive Samaritan's dilemma to hold.

Definition 2 The positive Samaritan's dilemma states that agent 0 can reach her first best when she moves in stage one and agents i , $i = 1, \dots, n$, move in stage two.

The game is solved by backwards induction. In stage two, each agent i , $i = 1, \dots, n$, sets the x_i that maximizes his own payoff, taking y_i and all other x_l , $l = 1, \dots, i - 1, i + 1, \dots, n$, as given:

$$\frac{\partial U_i}{\partial x_i} = 0. \quad (12)$$

In stage one, agent 0 sets the y_i that maximize her objective function $W(\mathbf{U})$, taking into account that agent i 's choice of x_i depend upon her choice of y_i :

$$W_0 \frac{\partial U_0}{\partial y_i} + W_i \frac{\partial U_i}{\partial y_i} + \sum_{k=0}^n W_k \frac{\partial U_k}{\partial x_i} \frac{dx_i}{dy_i} = 0.$$

Substituting (12) and differentiating it totally with respect to y_i , this becomes

$$W_0 \frac{\partial U_0}{\partial y_i} + W_i \frac{\partial U_i}{\partial y_i} - \frac{\partial^2 U_i / \partial y_i \partial x_i}{\partial^2 U_i / \partial x_i^2} \sum_{\substack{l=0 \\ l \neq i}}^n W_l \frac{\partial U_l}{\partial x_i} = 0. \quad (13)$$

In general, the outcome will not be agent 0's first best. We shall now see under which condition agent 0 can reach her first best when she moves first.⁷

Condition 1 For all \mathbf{x}^* as defined in Definition 1 and all $i = 1, \dots, n$,

$$\frac{\partial U_0}{\partial x_i} - \sum_{\substack{j=0 \\ j \neq i}}^n \frac{\partial U_0 / \partial y_j}{\partial U_j / \partial y_j} \frac{\partial U_j}{\partial x_i} = 0.$$

Proposition 1 Given that all agents' second order conditions are satisfied, the positive Samaritan's dilemma holds for all $W(\mathbf{U})$ if and only if Condition 1 holds.

The intuition behind the result is straightforward. When selfish agent i moves last, he does not take into account the effect of his action on any of the other agents' payoffs. This can only result in the first best for agent 0 if the net effect of agent i on other agents (weighted according to agent 0's objective function) is zero. Then agent i takes the full effect of his actions into account. There is no externality, and agent 0's first best is implemented.

In our introductory example of the Samaritan's dilemma (subsection 2.1), Condition 1 holds: the parasite's work effort does not affect the Samaritan's payoff. The Samaritan's own payoff only depends on his donation. In the introductory example of the rotten kid theorem (subsection 2.2), however, Condition 1 does not hold: the kid's action affects both his own and the parent's payoff.

⁷All proofs are available on the JEBO website.

3.3 Agents i move first

In this subsection, we derive the equilibrium for the game where agents i move before agent 0, and we see when this equilibrium corresponds to the first best for agent 0. Thus, we shall derive the conditions for the positive rotten kid theorem to hold.

Definition 3 *The positive rotten kid theorem states that agent 0 can reach her first best when agents i , $i = 1, \dots, n$, move in stage one and agent 0 moves in stage two.*

We solve the game by backwards induction. In stage two, agent 0 sets the y_j that maximize her objective function $W(\mathbf{U})$, taking all x_i , $i = 1, \dots, n$, as given:⁸

$$W_0 \frac{\partial U_0}{\partial y_j} + W_j \frac{\partial U_j}{\partial y_j} = 0. \quad (14)$$

In stage one, each agent i , $i = 1, \dots, n$, sets the x_i that maximizes his own payoff, taking the x_l , $l = 1, \dots, i-1, i+1, \dots, n$, from the other $n-1$ agents moving in stage one as given, but realizing that his choice of x_i affects agent 0's choice of y_i in stage two:

$$\frac{dU_i}{dx_i} \equiv \frac{\partial U_i}{\partial x_i} + \frac{\partial U_i}{\partial y_i} \frac{dy_i}{dx_i} = 0, \quad (15)$$

where the values for dy_j/dx_i , $j = 1, \dots, n$, follow from the total differentiation of (14) with respect to x_i .

In general, the equilibrium condition (15) for x_i , $i = 1, \dots, n$, is not identical to the corresponding first order condition (11) for agent 0's first best. We shall now see when it is.

⁸Obviously, these conditions are identical to the FOCs (10) for agent 0's first best with respect to \mathbf{y} .

Condition 2 Take a vector \mathbf{x}^* as defined in Definition 1. For this \mathbf{x}^* , write U_i as

$$U_i(y_i, \mathbf{x}) = G_i(\mathbf{x}) + z_i(y_i, \mathbf{x}) \quad (16)$$

with $\partial z_i / \partial y_i > 0$, $i = 1, \dots, n$. Then U_0 should satisfy

$$U_0(\mathbf{y}, \mathbf{x}) = G_0(\mathbf{x}) - F(\mathbf{z}) \quad (17)$$

with $\mathbf{z} \equiv (z_1, \dots, z_n)$, $\partial F / \partial z_i > 0$, and

$$\sum_{l=1}^n \frac{\partial^2 F}{\partial z_j \partial z_l} \frac{\partial G_l}{\partial x_i} = 0 \quad (18)$$

for all $i, j = 1, \dots, n$.

The sets of payoff functions that satisfy Condition 2 come in two categories:

1. $\partial^2 F / \partial z_j \partial z_l = 0$ for all $i, l = 1, \dots, n$ in (17). Then (16) and (17) become

$$U_0(y, \mathbf{x}) = G_0(\mathbf{x}) - \sum_{i=1}^n z_i(y_i, \mathbf{x}) \quad U_i(y_i, \mathbf{x}) = G_i(\mathbf{x}) + z_i(y_i, \mathbf{x}).$$

2. Not all $\partial^2 F / \partial z_j \partial z_l = 0$. Examples in this category are

$$\begin{aligned} U_0(y, \mathbf{x}) &= G_0(\mathbf{x}) - f_1(z_1) - z_2(y_2, \mathbf{x}) & f_1', f_1'' > 0 \\ U_i(y_i, \mathbf{x}) &= G_i(\mathbf{x}) + z_i(y_i, \mathbf{x}) & \frac{\partial G_1}{\partial x_i} = 0 \quad i = 1, 2 \end{aligned}$$

and

$$\begin{aligned} U_0(y, \mathbf{x}) &= G_0(\mathbf{x}) - e^{\alpha_1 z_1 + \alpha_2 z_2} \\ U_i(y_i, \mathbf{x}) &= \gamma_i G(\mathbf{x}) + z_i(y_i, \mathbf{x}) & \sum_i \alpha_i \gamma_i = 0 \quad i = 1, 2. \end{aligned}$$

Proposition 2 Given that all agents' second order conditions are satisfied, the positive rotten kid theorem holds for all $W(\mathbf{U})$ if and only if Condition 2 holds.

In order to interpret this result, let us state:

Lemma 1 *If and only if Condition 2 holds,*

1. *There is a single vector \mathbf{x}^* that implements the whole PPF.*

2. *The price of agent j 's payoff to agent 0 at $\mathbf{x} = \mathbf{x}^*$,*

$$P_j^* \equiv -\frac{\partial U_0(\mathbf{y}, \mathbf{x}^*)/\partial y_j}{\partial U_j(y_j, \mathbf{x}^*)/\partial y_j}, \quad (19)$$

is beyond manipulation by agent i :

$$\frac{dP_j^*}{dx_i} = 0. \quad (20)$$

Let us define a Utility Possibility Curve UPC as the set of vectors \mathbf{U} that can be obtained with a given \mathbf{x} . Lemma 1.1 says that the whole Payoff Possibility Frontier PPF must consist of a single UPC . Figure 1, inspired by Bergstrom's Figure 2, illustrates what goes wrong when a selfish agent can influence the price of his payoff or equivalently, when the PPF consists of multiple $UPCs$.

In Figure 1, point A on agent 0's indifference curve I_A is agent 0's first best. It is reached when the single selfish agent 1 selects the action x_A that implements UPC_A . When agent 1 cannot manipulate the price of his own payoff, all other $UPCs$ will be parallel and to the left of UPC_A . The whole PPF thus consists only of UPC_A . In that case, when U_1 is a normal good to the altruist, agent 1 will select x_A . However, suppose now that agent 1 can decrease the price of his own payoff, either by increasing or decreasing his x . For instance, when agent 1 chooses x_B , the resulting UPC_B is flatter than UPC_A , lies everywhere below I_A and intersects UPC_A so that the PPF does not consist of UPC_A alone. In point B , where agent 0's indifference curve I_B is tangent to UPC_B , U_1 is higher than in point A . Thus, agent 1 prefers implementing UPC_B to UPC_A .

When the selfish agents cannot influence the prices of their payoffs, we can aggregate all payoffs along the PPF for $\mathbf{x} = \mathbf{x}^*$ using these prices and refer to aggregate payoff as income $I(\mathbf{x})$. As Bergstrom (p. 1148) calls it, there is conditional transferable utility (conditional on \mathbf{x}). The agents i maximize income and agent 0 redistributes it. In the terminology of Monderer and Shapley (1996), Condition 2 turns the game into a potential game where all agents $i = 1, \dots, n$ maximize the ordinal potential function $I(\mathbf{x})$. Stated formally,

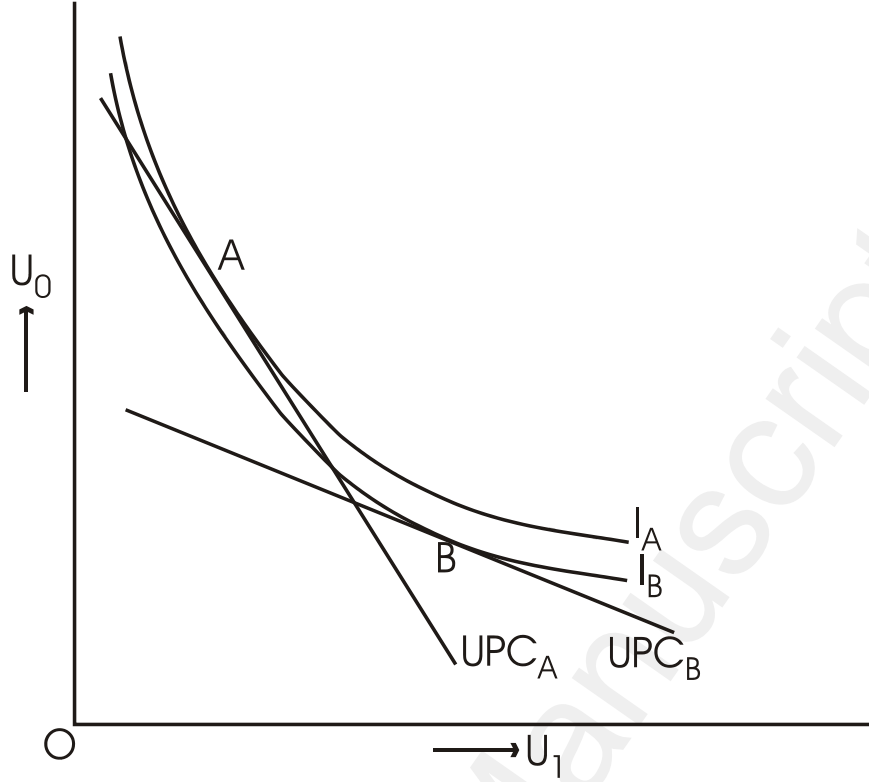


Figure 1: Intersecting Utility Possibility Curves

Lemma 2 *If and only if Condition 2 holds, we can define an income function*

$$I \equiv U_0(\mathbf{y}, \mathbf{x}) + \sum_{i=1}^n P_i^* U_i(y_i, \mathbf{x}) \quad (21)$$

(with P_i^* defined by (19)) that is a function of \mathbf{x} only. The payoff functions (16) and (17) can be written such that $I(\mathbf{x})$ is given by

$$I(\mathbf{x}) = \sum_{k=0}^n G_k(\mathbf{x}). \quad (22)$$

The income function $I(\mathbf{x})$ is maximized for $\mathbf{x} = \mathbf{x}^*$.

In our introductory example of the rotten kid theorem (subsection 2.2), Condition 2 holds because there is only one commodity, namely monetary income.⁹ Then we can identify payoffs with income and define aggregate or family income. All $UPCs$ are parallel and have slope -1 because they denote the feasible income distributions given aggregate

⁹Formally, the payoff functions (4) satisfy $z = y$ and $F(z) = z$, so that $\partial^2 F / \partial z^2 = 0$.

income from the kid's action. The kid cannot manipulate the price of his income to the parent because an extra dollar for the kid is always going to cost the parent one dollar.

In our example of the Samaritan's dilemma (subsection 2.1), however, there are two goods involved: money and the parasite's leisure. The parasite can manipulate the price of his payoff to the Samaritan by his choice of leisure. By working less, the parasite has less money of his own and a higher marginal payoff of money. In this way he lowers the price of his payoff to the Samaritan, so that the Samaritan will buy more of it.

4 Bergstrom's rotten kid game

The present paper is not the first to have derived conditions for the rotten kid theorem to hold. Bergstrom (1989) and Cornes and Silva (1999) have previously derived a condition from a model more specific than ours. In their model, the altruist distributes a certain sum of money among the selfish agents. The total amount of money available may depend on the selfish agents' actions.

In subsection 4.1, we introduce Bergstrom's game and his own sufficient condition for the positive rotten kid theorem. We shall see that as his maximization problem for the altruist is a special case of our more general problem, his payoff condition is accordingly a special version of our payoff condition. We shall also find that Bergstrom was wrong in claiming that the payoff condition is necessary only when "money is important enough". In subsection 4.2, we discuss Cornes and Silva's condition for the positive rotten kid theorem to hold in Bergstrom's model. We shall see that this condition does not carry over to our own more general model and that there are no further solutions to our or Bergstrom's model.

4.1 Bergstrom's solution

In Bergstrom's model, the role of the altruist is limited to the distribution of a certain amount of money. There are three steps involved in moving from our model to Bergstrom's. First, agent 0's actions \mathbf{y} are restricted to giving money to the selfish agents. The relevant property of money in this context is the following:

Definition 4 When \mathbf{y} is money, agent 0's payoff depends on how much she does on aggregate for all other agents, but not on the distribution of this total amount among the agents. Then the altruist's payoff is given by $U_0(y_0, \mathbf{x})$ with $y_0 \equiv \sum_{i=1}^n y_i$.

When \mathbf{y} is money, $\partial U_0/\partial y_i = \partial U_0/\partial y_0$ for all $i = 1, \dots, n$. Applying this to Condition 2, we see that the payoff functions should satisfy

$$U_0(y, \mathbf{x}) = G_0(\mathbf{x}) - H[A(\mathbf{x})y_0] \quad (23)$$

$$U_i(\mathbf{x}, y_i) = G_i(\mathbf{x}) + A(\mathbf{x})y_i \quad (24)$$

for $i = 1, \dots, n$, with $H' > 0$ and

$$H''[A(\mathbf{x})y_0] \sum_{j=1}^n \frac{\partial G_j(x)}{\partial x_i} = 0. \quad (25)$$

The second step from our framework to Bergstrom's is that the budget constraint $\sum_{i=1}^n y_i \leq y(\mathbf{x})$ is binding. Then the functional form of $U_0(y, \mathbf{x})$ is irrelevant, so that (23) and (25) are no longer needed. Condition (24) must still hold because the equivalent of condition (20) is now

$$d \left(\frac{\partial U_l(y_l, \mathbf{x})/\partial y_l}{\partial U_j(y_j, \mathbf{x})/\partial y_j} \right) / dx_i = 0$$

for all $i, j, l = 1, \dots, n$. The third and final step from our framework to Bergstrom's is to exclude U_0 from agent 0's objective function. This final step does not lead to additional constraints on the payoff functions.

Proposition 3 Given that the second order conditions are satisfied, the positive rotten kid holds for all $W(\mathbf{U})$ and all $y(\mathbf{x})$ in agent 0's maximization problem

$$\max W(U_1(y_1, \mathbf{x}), \dots, U_n(y_n, \mathbf{x})) \quad s.t. \quad \sum_{i=1}^n y_i = y(\mathbf{x}) \quad (26)$$

if and only if all U_i , $i = 1, \dots, n$, have the form (24).

Condition (24) is identical to Bergstrom's payoff condition. Our Condition 2 is more general than Bergstrom's as illustrated by the fact that none of our exemplary payoff functions given in subsection 3.3 satisfy condition (24). The reason why Bergstrom's condition is more restrictive than ours is that he restricts the altruist's actions to the

distribution of a certain amount of money. We shall now explore the intuition behind this result.

As we know from Corollary 1.1, the positive rotten kid theorem holds when there is only one vector \mathbf{x}^* , or equivalently, one Utility Possibility Curve, that implements the whole Payoff Possibility Frontier. When \mathbf{y} is money, an agent's payoff on a *UPC* depends only on how much money he gets. That means we can identify an agent's payoff with the amount of money he gets. This implies a one-to-one tradeoff between all agents' payoffs on the whole *UPC* and thereby on the whole *PPF*. Thus, when \mathbf{y} is money, the prices of payoffs are constant along the *PPF*. However, this is not a necessary condition for the positive rotten kid theorem. Prices can vary along the *PPF* with the altruist's actions \mathbf{y} , as long as they cannot be manipulated by the selfish agents' actions \mathbf{x} . To put it differently, payoff prices can vary as long as there is a single vector of \mathbf{x} that maximizes total payoff, at whatever prices payoffs are aggregated.

Our Proposition 3 states that condition (24) is necessary and sufficient for the positive rotten kid theorem to hold. Bergstrom, however, claims that the condition is sufficient, but only necessary when combined with two further conditions. These are that all U_i are normal goods and that money is important enough. We have already mentioned the normal good assumption in subsections 2.2 and 3.3. It can be shown that this assumption is necessary and sufficient for the second order conditions to hold. In our analysis, we have simply assumed that the second order conditions hold. However, we have not encountered anything resembling the condition that money is important enough. We shall now see that indeed this condition is redundant.

In the terminology of our paper, Bergstrom's condition that money is important enough can be stated as follows:

Condition 3 $\partial U_i(y_i, \mathbf{x})/\partial y_i > 0$ for all $y_i > 0$, $i = 1, \dots, n$, and for all feasible \mathbf{x} . There is some vector of actions \mathbf{x}^0 such that for every agent i , all y_i , and all feasible \mathbf{x} , there exists y'_i such that $U_i(y'_i, \mathbf{x}^0) = U_i(y_i, \mathbf{x})$.

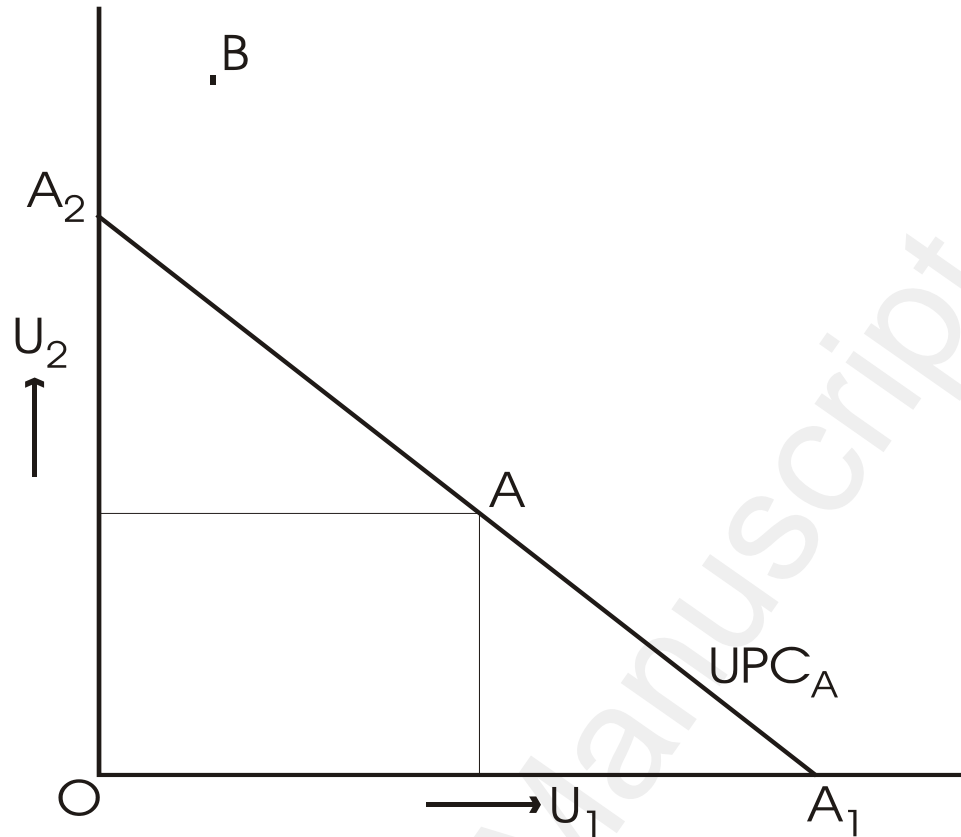


Figure 2: When money is not important enough

Bergstrom uses this condition to show that there is always a Utility Possibility Curve with slope -1 . That is there is a vector \mathbf{x}^0 such that:

$$\frac{\partial U_i(y_i, \mathbf{x}^0)/\partial y_i}{\partial U_j(y_j, \mathbf{x}^0)/\partial y_j} = -1,$$

for all y_i, y_j and for all $i, j = 1, \dots, n$. However, all that is needed to prove this is the first part of Condition 3 which we also used in our analysis, that $\partial U_i/\partial y_i > 0$. As we have argued above, given any vector \mathbf{x}^0 that implements a point on the *PPF*, a kid i 's payoff depends only on the amount of money he gets from the parent. This means we can identify agent i 's payoff given this vector \mathbf{x}^0 with the amount of money he gets.

Another way of looking at the issue is illustrated with Figure 2 with two selfish agents, 1 and 2. Point A is on the Payoff Possibility Frontier *PPF* and on Utility Possibility Curve UPC_A , implemented by (x_1^A, x_2^A) . The maximum utility for agent 2 on UPC_A is A_2 . If money were not important enough, there would be another vector \mathbf{x} with which U_2 could

exceed A_2 and a point like B on UPC_B (UPC_B is not shown in Figure 2) would be feasible. However, since point A is on the PPF , UPC_B would have to be steeper than UPC_A and would have to cross it at some point. As we have seen with Figure 1, the rotten kid theorem does not hold when the PPF consists of multiple UPC s with different slopes, crossing each other. Thus, when the rotten kid theorem holds, point B is not feasible with any \mathbf{x} . We conclude:

Lemma 3 *When the rotten kid theorem holds for all $W(\mathbf{U})$ and all $y(\mathbf{x})$ in agent 0's maximization problem (26), money is important enough.*

4.2 Cornes and Silva's solution

Cornes and Silva recently found another condition for the positive rotten kid theorem to hold in Bergstrom's framework. Under this condition, all kids contribute to a pure public good. In this subsection we shall first discuss Cornes and Silva's result in the light of our own analysis, demonstrating why it does not carry over to our more general framework. We shall also argue that there are no additional conditions under which the rotten kid theorem holds for all $W(\mathbf{U})$, neither in Bergstrom's framework, nor in our more general setup. Finally, we shall discuss the problems associated with Cornes and Silva's solution and conclude that they do not carry over to our general framework, if interpreted strictly.

In the notation of this paper, Cornes and Silva's model can be described as follows. Agent i , $i = 1, \dots, n$, initially has an exogenous endowment m_i . From this endowment he can make a contribution x_i to the public good $X \equiv \sum_{i=1}^n x_i$. The rest of his endowment plus the transfer t_i from agent 0 is available for consumption y_i of the private good. Agent 0 has no budget of her own: $\sum_{i=1}^n t_i = 0$. Agent 0's budget constraint can also be written as $\sum_{i=1}^n y_i = M - X$, with $M \equiv \sum_{i=1}^n m_i$.

How did Cornes and Silva manage to find this additional solution? To find that out, let us first briefly present the derivation of Bergstrom's own solution with our method for deriving Proposition 2. Adapting equation (30) from the proof of Proposition 2, we find that $dU_j/dx_i = 0$ must hold for all $i, j = 1, \dots, n$ for the rotten kid theorem to apply for all $W(\mathbf{U})$ and all $y(\mathbf{x})$. The agents i set $dU_i/dx_i = 0$ themselves. We need conditions on \mathbf{U} to make sure that agent 0 will set $dU_l/dx_i = 0$ for all other $l, i = 1, \dots, n, l \neq i$. These

conditions are (24).

Instead of having agent 0 set all $dU_l/dx_i = 0$ herself, we could impose some restrictions R on the payoff functions so that $dU_i/dx_i = 0$ automatically implies $dU_l/dx_i = 0$ for some (but not all) $l, i = 1, \dots, n, l \neq i$. However, it can be shown that as long as agent 0 still has to set some $dU_l/dx_i = 0$ herself, the payoff condition will simply be (24) with restrictions R .

The only option left is then to impose that when agent i sets $dU_i/dx_i = 0$, this should automatically imply $dU_l/dx_i = 0$ for all $l, i = 1, \dots, n, l \neq i$. This will be the case if and only if we can define $X \equiv \sum_{i=1}^n x_i$. Then the payoff functions become $U_i(y_i, \mathbf{x}) = U_i(y_i, X)$, and the resource constraint turns into $y(\mathbf{x}) = y(X)$. The n^2 conditions $dU_i/dx_j = 0$ for implementation of agent 0's first best reduce to n conditions $dU_i/dX = 0$. Agents i 's first order conditions are also $dU_i/dX = 0$. Without loss of generality, we can specify $y(X) = M - X$. Then we have reproduced Cornes and Silva's pure public good case. Note that Bergstrom's result, as stated here in Proposition 3, still stands because Cornes and Silva introduce a restriction $y(\mathbf{x}) = y(X)$ on the resource constraint. If we allow for restrictions on $y(\mathbf{x})$, then the only additional payoff condition for the rotten kid theorem is Cornes and Silva's.

We can now see why Cornes and Silva's condition does not carry over to our more general framework. When $X \equiv \sum_{i=1}^n x_i$, the agents $i, i = 1, \dots, n$, will set $dU_i/dX = 0$. However, we still have to make sure that agent 0 will set $dU_0/dX = 0$. She will do this if and only if the payoff functions satisfy Condition 2 with \mathbf{x} replaced by X . Thus, it is impossible to find any solution other than Condition 2 in the general framework.

Two problems have been noted with regard to this solution. Cornes and Silva acknowledge that there are multiple equilibria because only the equilibrium (and optimum) amount of the public good X is determined, but individual contributions x_i are not. All interior equilibria implement the optimum. However, Chiappori and Werning (2002) note that in general, there is no interior solution to the game. Both problems have the same root cause: There is only one optimum condition for X , whereas there are n equilibrium conditions for \mathbf{x} in the game. Either the conditions for an interior equilibrium are compatible with each other, in which case there are multiple equilibria (Cornes and Silva), or

they are not, in which case there is no interior solution (Chiappori and Werning).

Strictly speaking, the Cornes and Silva solution is not admissible in our framework because we have assumed in subsection 3.1 that the altruist's first best is unique in (\mathbf{y}, \mathbf{x}) . Since the problems of multiple equilibria and nonexistence of interior solutions discussed here derive from the non-uniqueness of the altruist's first best, we can be assured that these problems, in this form, will not occur in our framework.

5 Conclusion

For thirty years, Buchanan's (1975) Samaritan's dilemma and Becker's (1974) rotten kid theorem, with their mutually exclusive claims, have coexisted in the economic theory of altruism. This paper has been the first to analyze the conditions on the payoff functions under which either result holds for any altruistic objective function. We have seen that the altruist can reach her first best when she moves first if and only if a selfish agent's action does not on balance affect any other agent's payoff in the optimum. Then there are no externalities to the selfish agents' actions. The altruist can reach her first best when she moves last if and only if the selfish agents cannot manipulate the altruist's trade-off between her own and the selfish agents' payoffs. Then the selfish agents will maximize aggregate payoff and the altruist will redistribute income.

The focus of this paper has been on the simple one-shot game with complete information with which the theory started in the mid-1970s. Since then, more complex games between altruists and selfish agents have been studied.¹⁰ It would be worthwhile to expand the general analysis to encompass multi-period models and incomplete or asymmetric information. The former is especially relevant as we would expect to find altruism mainly in ongoing relations.

The theory of altruism can also be applied to government policy. The link between these two fields of research is that the government can be regarded as an altruist when it maximizes social welfare or any other objective function that depends positively on the payoff of other players. Thus, the theory of altruism can contribute to our understanding

¹⁰Among others, Lindbeck and Weibull (1988), Bruce and Waldman (1990, 1991) and Futagami et al. (2004) have analyzed two-period lifetime models. Wirl (1999) and Lagerlöf (2004) assume asymmetric information. Coate (1995) and Lord and Raganzas (1995) include uncertainty.

of when collective and individual interests coincide (Shapiro and Petchey 1998, Munger 2000). Under the conditions of the Samaritan's dilemma, the government can reach the optimum if and only if it can commit to a certain policy. If the Samaritan's dilemma does not apply, commitment does not result in the first best. The government may then be better off with a time-consistent policy. Under the conditions of the rotten kid theorem, time-consistent policy even results in the first best. Starting with Kydland and Prescott (1977), most analyses of time consistency have used a more complicated setup than ours. We offer a general framework, akin to Chari et al. (1989) and Cubitt (1992), along with results to be applied to games between the government and private agents.

6 Appendix

Proof of Proposition 1. The necessary and sufficient condition for (12) to turn into (11) and for (13) to turn into (10) is

$$\sum_{\substack{l=0 \\ l \neq i}}^n W_l \frac{\partial U_l}{\partial x_i} = 0 \quad (27)$$

for all $i = 1, \dots, n$. Substituting (27) and (12) into (13) yields Condition 1.

Proof of Proposition 2. Since agent 0 moves last, the first order conditions (10) for agent 0's first best with respect to \mathbf{y} are satisfied. Substituting (10) and (15), we can rewrite the first best conditions (11) for \mathbf{x} as

$$\sum_{\substack{l=0 \\ l \neq i}}^n W_l \frac{dU_l}{dx_i} = W_0 \left(\frac{dU_0}{dx_i} - \sum_{\substack{j=1 \\ j \neq i}}^n \frac{\partial U_0 / \partial y_j}{\partial U_j / \partial y_j} \frac{dU_j}{dx_i} \right) = 0 \quad (28)$$

for all $i = 1, \dots, n$, where dU_k/dx_i , $k = 0, \dots, n$, is defined by

$$\begin{aligned} \frac{dU_0}{dx_i} &\equiv \frac{\partial U_0}{\partial x_i} + \sum_{j=1}^n \frac{\partial U_0}{\partial y_j} \frac{dy_j}{dx_i} \\ \frac{dU_j}{dx_i} &\equiv \frac{\partial U_j}{\partial x_i} + \frac{\partial U_j}{\partial y_j} \frac{dy_j}{dx_i}. \end{aligned}$$

To find the expressions for dU_k/dx_i , write the total differential of agent 0's first order

condition (14) with respect to x_i , using (10), as

$$\left[\frac{d(\partial U_0/\partial y_j)}{dx_i} - \frac{\partial U_0/\partial y_j}{\partial U_j/\partial y_j} \frac{d(\partial U_j/\partial y_j)}{dx_i} \right] + W_0 \left[\sum_{\substack{l=0 \\ l \neq i}}^n \left(W_{jl} \frac{\partial U_j}{\partial y_j} + W_{0l} \frac{\partial U_0}{\partial y_j} \right) \frac{dU_l}{dx_i} \right] = 0. \quad (29)$$

We want to obtain solutions for dU_k/dx_i that don't contain second derivatives of W , because we don't want to put any restrictions on these. The only way to do this is by setting both terms between large square brackets on the LHS of (29) equal to zero. With the second term equal to zero, there can only be definite solutions to dU_l/dx_i , $i = 1, \dots, n$, $l = 0, \dots, n$, $l \neq i$, if

$$\frac{dU_k}{dx_i} = 0 \quad \text{for all } k = 0, \dots, n, i = 1, \dots, n \quad (30)$$

where $dU_i/dx_i = 0$ by (15). When (30) holds, (28) is satisfied and all conditions for the implementation of agent 0's first best are met.

We can rewrite the condition that the first term in large square brackets on the LHS of (29) equals zero as

$$d \left(\frac{\partial U_0/\partial y_j}{\partial U_j/\partial y_j} \right) / dx_i = 0. \quad (31)$$

We can always write the agents' payoffs as

$$U_0(\mathbf{y}, \mathbf{x}) = g_0(\mathbf{x}) - H(\mathbf{v}, \mathbf{x}) \quad (32)$$

$$U_i(y_i, \mathbf{x}) = g_i(\mathbf{x}) + v_i(y_i, \mathbf{x}) \quad (33)$$

with $\mathbf{v} \equiv (v_1, \dots, v_n)$, $\partial H/\partial v_i > 0$, $\partial v_i/\partial y_i > 0$, $i = 1, \dots, n$. Replacing y_j by v_j and substituting (32) and (33), condition (31) becomes

$$\frac{d(\partial H/\partial v_j)}{dx_i} = \frac{\partial^2 H}{\partial v_j \partial x_i} + \sum_{l=1}^n \frac{\partial^2 H}{\partial v_j \partial v_l} \frac{dv_l}{dx_i} = \frac{\partial^2 H}{\partial v_j \partial x_i} - \sum_{l=1}^n \frac{\partial^2 H}{\partial v_j \partial v_l} \frac{\partial g_l}{\partial x_i} = 0. \quad (34)$$

The second equality follows from (30). The following lemma completes the proof:

Lemma 4 *Equations (32), (33) and (34) can always be written as (17), (16) and (18).*

Proof. The result is obvious if $\partial^2 H/\partial v_j \partial x_i = 0$ for all $i, j = 1, \dots, n$. Now suppose there is a $\partial^2 H/\partial v_j \partial x_i \neq 0$. Then, for the \mathbf{v} terms to drop out of (34), the function $H(\mathbf{v}, \mathbf{x})$ in (32) must have the form

$$H(\mathbf{v}, \mathbf{x}) = e^{\sum_{i=1}^n \alpha_i v_i} h(\mathbf{x}) + \sum_{i=1}^n \beta_i v_i$$

with $\alpha_i, \beta_i \geq 0$. Then

$$\frac{\partial^2 H(\mathbf{v}, \mathbf{x})}{\partial v_j \partial x_i} = \alpha_j e^{\sum_i \alpha_i v_i} \frac{\partial h(\mathbf{x})}{\partial x_i} \quad \frac{\partial^2 H(\mathbf{v}, \mathbf{x})}{\partial v_j \partial v_l} = \alpha_j \alpha_l e^{\sum_i \alpha_i v_i} h(\mathbf{x})$$

so that by (34),

$$\sum_{l=1}^n \alpha_l \frac{\partial g_l(\mathbf{x})}{\partial x_i} = \frac{\partial h / \partial x_i}{h(\mathbf{x})}.$$

Then for all l with $\alpha_l > 0$, $g_l(\mathbf{x})$ must have the form

$$g_l(\mathbf{x}) = \gamma_l \ln h(\mathbf{x}) + G_l(\mathbf{x})$$

with $\sum_l \alpha_l \gamma_l = 1$ and $\sum_l \alpha_l \partial G_l / \partial x_i = 0$.

Then we can define z_i as

$$z_i(y_i, \mathbf{x}) \equiv v_i(y_i, \mathbf{x}) + \gamma_i \ln h(\mathbf{x})$$

so that U_0 and U_i have the form (17) and (16) respectively, with

$$\begin{aligned} F(\mathbf{z}) &= e^{\sum_i \alpha_i z_i} + \sum_{i=1}^n \beta_i z_i \\ G_0(\mathbf{x}) &= g_0(\mathbf{x}) + \sum_{i=1}^n \beta_i \gamma_i \ln h(\mathbf{x}) \quad G_i(\mathbf{x}) = g_i(\mathbf{x}) + \gamma_i \ln h(\mathbf{x}). \end{aligned}$$

Finally, with $\partial^2 F / \partial z_j \partial x_i = 0$, (34) turns into (18). ■

Proof of Lemma 1.1. Equation (31) implies that Utility Possibility Curves (*UPCs*) cannot cross each other. Then either the whole Payoff Possibility Frontier consists of a single *UPC*, or there are layers of *UPCs* with the *PPF* tracing their outlines, as in Figure 3. A_0A_1 and B_0B_1 are two members of a family of parallel *UPCs* shrinking to a single point at C . The *PPF* is given by V_0V_1 . In agent 0's first best (except if it is at point C), her indifference curve is not tangent to the *UPC*. In Figure 3, for instance, agent 0's optimum is at B_1 on indifference curve I_B . B_1 is a corner solution: Agent 0 would like to give more to agent 1, but she has already given him all she has got. First order condition (10) does not hold. Since we have assumed agent 0's first best is an interior solution, we cannot allow for a *PPF* tracing the outlines of parallel *UPCs*. Thus, the whole *PPF* must consist of a single *UPC*.

2. Equation (20) is equation (31) from the proof of Proposition 2.

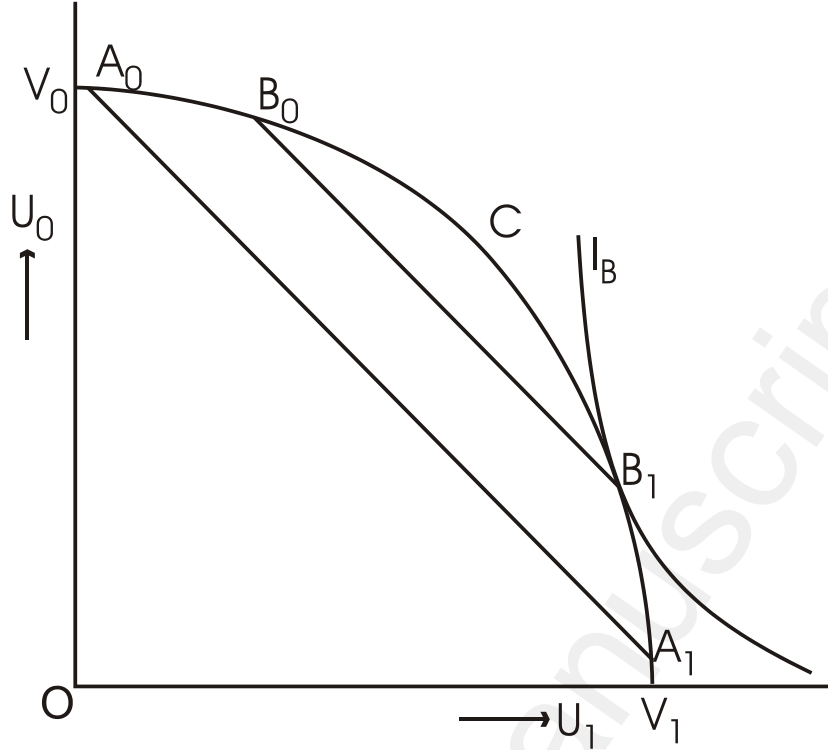


Figure 3: A Payoff Possibility Frontier tracing the outlines of parallel Utility Possibility Curves

Proof of Lemma 2. Applying agent 0's first best conditions for y_i (10) and x_j (11) to the payoff functions of Condition 2 yields, respectively,

$$W_0 \frac{\partial F}{\partial z_i} = W_i \quad (35)$$

$$W_0 \left(\frac{\partial G_0}{\partial x_j} - \frac{\partial F}{\partial z_i} \frac{\partial z_i}{\partial x_j} \right) + \sum_{i=1}^n W_i \left(\frac{\partial G_i}{\partial x_j} + \frac{\partial z_i}{\partial x_j} \right) = 0. \quad (36)$$

Substituting (35) into (36),

$$\frac{\partial G_0}{\partial x_j} + \sum_{i=1}^n \frac{\partial F}{\partial z_i} \frac{\partial G_i}{\partial x_j} = 0. \quad (37)$$

This implies that \mathbf{x}^* maximizes income $I(\mathbf{x})$, as given in (22), if and only if all $\partial F/\partial z_i$ can be replaced by ones. We shall now see how this can be accomplished.

1. When $\partial F/\partial z_i$ is a constant, we can normalize it to one.
2. When $\partial F/\partial z_i$ is not a constant, then $\partial^2 F/\partial z_i \partial z_l \neq 0$ for some $l, l = 1, \dots, n$. If

there is an i with only one $\partial^2 F/\partial z_i \partial z_l \neq 0$, then (18) implies that $\partial G_l/\partial x_j = 0$. Then we can set $\partial F/\partial z_l$ equal to any expression, including $\partial F/\partial z_l = 1$.

3. For those i with more than one $\partial^2 F/\partial z_i \partial z_l \neq 0$, we substitute $\partial G_l/\partial x_j = 0$ for those l identified in step 2 into (18). If this leaves only one term,

$$\frac{\partial^2 F}{\partial z_j \partial z_l} \frac{\partial G_l}{\partial x_j} = 0,$$

with $\partial^2 F/\partial z_i \partial z_l \neq 0$, then obviously $\partial G_l/\partial x_j = 0$ for this l , and we can set $\partial F/\partial z_l = 1$.

1. Substitute $\partial G_l/\partial x_j = 0$ into the expressions (18) for the remaining i , and so on.

4. If there are still i left with more than one term in their expression (18), then we can write this expression as

$$\Gamma_i(\mathbf{z}) \sum_{l=1}^n \alpha_{il} \frac{\partial G_l}{\partial x_j} = 0. \quad (38)$$

This is because F is a function of \mathbf{z} only and G is a function of \mathbf{x} only.

- (a) If there is only one i left, then rescaling all G_l functions in (38) such that all α_l are normalized to one yields

$$\sum_{l=1}^n \frac{\partial G_l}{\partial x_j} = 0,$$

and we can set $\partial F/\partial z_l = 1$ for all l involved.

- (b) If there is more than one i left, then the unique solution to the system of (18) equations for these i is $\partial G_l/\partial x_j = 0$ for all l involved. Again, we can set $\partial F/\partial z_l = 1$ for all l involved.

References

- Becker, G.S., 1974. A theory of social interaction. *Journal of Political Economy* 82, 1063-1093.
- Becker, G.S., 1976. Altruism, egoism, and genetic fitness: Economics and sociobiology. *Journal of Economic Literature* 14, 817-826.
- Bergstrom, T.C., 1989. A fresh look at the rotten kid theorem— and other household mysteries. *Journal of Political Economy* 97, 1138-1159.

Bernheim, B.D., Shleifer, A., Summers, L.H., 1985. The strategic bequest motive. *Journal of Political Economy* 93, 1045-1076.

Bruce, N., Waldman, M., 1990. The rotten-kid theorem meets the Samaritan's dilemma. *Quarterly Journal of Economics* 105, 155-165.

Bruce, N., Waldman, M., 1991. Transfers in kind: Why they can be efficient and nonpaternalistic. *American Economic Review* 81, 1345-1351.

Buchanan, J.M., 1975. The Samaritan's dilemma. In: Phelps, E.S. (Ed.). *Altruism, Morality, and Economic Theory*. New York: Sage Foundation, 71-85.

Chari, V.V., Kehoe, P.J., Prescott, E.C., 1989. Time consistency and policy. In: Barro, R.J. (Ed.). *Modern Business Cycle Theory*. Cambridge (Mass.): Harvard University Press, 265-305.

Chiappori, P-A., Werning, I., 2002. Comment on "Rotten kids, purity, and perfection". *Journal of Political Economy* 110, 475-480.

Cremer, H., Pestieau, P., 1996. Bequests as a heir "discipline device". *Journal of Population Economics* 9, 405-416.

Coate, S., 1995. Altruism, the Samaritan's dilemma, and government transfer policy. *American Economic Review* 85, 46-57.

Cornes, R.C, Silva, E.C.D., 1999. Rotten kids, purity, and perfection. *Journal of Political Economy* 107, 1034-1040.

Cubitt, R.P., 1992. Economic policy precommitment and social welfare. *Journal of Public Economics* 49, 191-201.

Futagami, R., Kamada, K., Sato, T., 2004. Government transfers and the Samaritan's dilemma in the family. *Public Choice* 118, 77-86.

Henrich, J., 2004. Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior and Organization* 53, 3-35.

Hirshleifer, J., 1977. Shakespeare vs. Becker on altruism: The importance of having the last word. *Journal of Economic Literature* 15, 500-502.

Jürges, H., 2000. Of rotten kids and Rawlsian parents: The optimal timing of intergenerational transfers. *Journal of Population Economics* 13, 147-157.

Kydland, F.E., Prescott, E.C., 1977. Rules rather than discretion: The inconsistency

of optimal plans. *Journal of Political Economy* 85, 473-491.

Lagerlöf, J., 2004. Efficiency-enhancing signalling in the Samaritan's dilemma. *Economic Journal* 114, 55-68.

Lindbeck, A., Weibull, J.W., 1988. Altruism and time consistency: The economics of fait accompli. *Journal of Political Economy* 96, 1165-1182.

Lord, W., Raganzas, P., 1995. Uncertainty, altruism, and savings: Precautionary savings meets the Samaritan's dilemma. *Public Finance* 50, 404-419.

Monderer, D., Shapley, L.S., 1996. Potential games. *Games and Economic Behavior* 14, 124-143.

Munger, M.C., 2000. Five questions: An integrated research agenda for Public Choice. *Public Choice* 103, 1-12.

Schmidtchen, D., 1999. To help or not to help: The Samaritan's dilemma revisited. Discussion paper 9909, Center for the Study of Law and Economics, Saarland University.

Shapiro, P., Petchey, J., 1998. The coincidence of collective and individual interests. Working paper 9-98R, Department of Economics, University of California, Santa Barbara.

Wirl, F., 1999. Paternalistic principals, *Journal of Economic Behavior and Organization* 38, 403-419.