UNIVERSITY OF NOTTINGHAM

Discussion Papers in Economics

Discussion Paper
No. 02/13

SAMARITAN VS. ROTTEN KID: ANOTHER LOOK

by Bouwe R. Dijkstra

November 2002

DP 02/13

ISSN 1360-2438

UNIVERSITY OF NOTTINGHAM

Discussion Papers in Economics

Discussion Paper
No. 02/13

SAMARITAN VS. ROTTEN KID: ANOTHER LOOK

by Bouwe R. Dijkstra

Bouwe Dijkstra is Lecturer, School of Economics, University of Nottingham

November 2002

Samaritan vs Rotten Kid: Another Look

Bouwe R. Dijkstra*

University of Nottingham

October 2002

Abstract. We set up a two-stage game with sequential moves by one altruistic agent and n selfish agents. The rotten kid theorem states that the altruist can only reach her first best when the selfish agents move before the altruist. The Samaritan's dilemma, on the other hand, states that the altruist can only reach her first best when she moves before the selfish agents. 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 only affects 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 payoff. When the altruist cannot reach her first best when she moves last, the outcome is not Pareto efficient either.

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, Johan Lagerlöf, Andreas Lange, Markus Lehmann, Andries Nentjes, Till Requate, Bert Schoonbeek, Perry Shapiro and Claudio Zoli for valuable comments. Special thanks are due to Ted Bergstrom for essential clarifications.

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 altruism offers two conflicting pieces of strategic advice: the rotten kid theorem (Becker [1] [2]) and the Samaritan's dilemma (Buchanan [7]). 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 only when the Samaritan moves first will his actions be immune to manipulation by the parasite.

In this paper, we shall 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 only affect his own payoff on the margin. 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.

As we interpret Samaritan's dilemma and rotten kid theorem, they have a positive as well as a negative side. The positive side is that the altruist can reach her first best under one sequence of moves. The negative side is that she cannot reach her first best under the other sequence. Many authors have used the terms Samaritan's dilemma and rotten kid theorem in the positive sense only. We shall refer to these versions as the positive Samaritan's dilemma and the positive rotten kid theorem.

Our result for the positive Samaritan's dilemma is new. For the positive rotten kid theorem, Bergstrom [3] has performed a similar analysis. His model can be seen as a special version of our more general setup. Whereas we do not restrict the nature of the altruist's actions, Bergstrom [3] assumes she distributes a certain amount of money among the selfish agents. Removing this restriction results in a more general condition for the positive rotten kid theorem. Bergstrom [3] also claims 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 [12] have found another condition for the rotten kid theorem to hold in Bergstrom's [3] framework. We shall see that this condition only applies in Bergstrom's [3] 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 [23]). 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 (Dijkstra [15]).

The rest of this paper is organized as follows. In Section 2, we introduce and discuss 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 Samaritan's dilemma and rotten kid theorem in terms of Pareto efficiency. In Section 5, we discuss Bergstrom's [3] game as well as Bergstrom's [3] own and Cornes and Silva's [12] conditions for the rotten kid theorem. We conclude with Section 6.

2 Introductory examples

2.1 Samaritan's dilemma

The Samaritan's dilemma is due to Buchanan [7] who discusses a game between an altruistic Samaritan and a selfish parasite.[1] He shows that the Samaritan can reach his first

[1] Buchanan [7] distinguishes between the active and the passive Samaritan's dilemma. We shall only discuss the passive Samaritan's dilemma here. The passive Samaritan's preferences are reconcilable with

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.[2]

The Samaritan maximizes his objective function $W(U_0, U_1)$, 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 only depends 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:

- $\partial U_1 / dy > 0$, $\partial^2 U_1 / \partial y^2 \leq 0$. The parasite's marginal payoff of money is positive and decreasing.

- $\partial U_1 / \partial x > [<] 0$ for $x < [>] x^*(y), x^*(y) > 0$. 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 more money when he works harder 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.

a 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 follows from the fact that, given that the Samaritan donates, he prefers the parasite to go to work although the parasite prefers to stay in bed. Schmidtchen [25] provides an analysis of the active Samaritan's dilemma.

[2] Jürges [17] also analyzes this game. Bergstrom ([3], 1140-1) analyzes a similar game, where a parent distributes money after his "lazy rotten kids" have set their work efforts. Neither Bergstrom [3] nor Jürges [17] identify the game with the Samaritan's dilemma.

When the Samaritan moves 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). This means that 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 only affects 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 only corresponds 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):

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

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 more 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 [1] [2] and commented upon by Hirshleifer [16]. The game is between an altruistic parent and a selfish kid. The kid can undertake 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.

In fact, Becker [1] [2] himself does not discuss the order of moves. Citing Shakespeare's *King Lear*, Hirshleifer [16] was the first to point out that the parent's first best is implemented only when the kid moves first.[3]

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:

$$U_0 = -y + b_0(x) \qquad U_1 = y + b_1(x) \tag{4}$$

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 respectively:

$$W_0 = W_1 \tag{5}$$

$$W_0 b_0' + W_1 b_1' = 0 \tag{6}$$

Substituting (5) into (6):

$$b_0' + b_1' = 0 \tag{7}$$

[3] Pollak [24] offers an alternative qualification: The parent can reach her first best only if she makes a take-it-or-leave-it offer to the kid. The offer specifies the kid's action and the parent's transfer. Cox [13] elaborates on this point. He argues that the parent can only reach her first best if the kid is better off accepting the offer to implement the first best than rejecting it. Cox [13] calls this "altruism". If the kid's participation constraint is binding, the parent will offer a different contract which gives the kid his reservation payoff. Cox calls this "exchange". In our model, we assume that the selfish agent's participation constraint never binds.

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 shall see what happens when the kid moves 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 income, taking into account that his action affects the parent's transfer:

$$\frac{dU_1}{dx} \equiv \frac{dy}{dx} + b_1' = 0 \tag{8}$$

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

$$(W_{00} - W_{10})\left(-\frac{dy}{dx} + b_0'\right) = (W_{11} - W_{01})\left(\frac{dy}{dx} + b_1'\right) \tag{9}$$

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. As Bernheim et al. [4] already noted, this result follows from the assumption that there is only one commodity, namely income. The intuition, due to Bergstrom [3], 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 [1] already notes, the kid will only want to maximize family income if he benefits from that himself, i.e. 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 $k = 0, \cdots, n$. Agent 0 is the altruist and agents $i,\ i = 1, \cdots, n$, 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 $i, j = 1, \cdots, n,\ i \neq j$, by definition.

The vector $\mathbf{y} = (y_1, \cdots, y_n)$ has to be feasible. There is a lower bound of $\mathbf{y} = \mathbf{0}$: agent 0 can only give to the other agents, she cannot improve her own payoff at the expense of the others. There is also an upper bound to $\mathbf{y}$, which follows from the restriction that agent 0 only has 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 shall assume that neither the upper nor the lower bound are binding constraints on the equilibria.

Agent 0's payoff has the form $U_0(\mathbf{y}, \mathbf{x})$, which is continuous and twice differentiable, with $\mathbf{x} = (x_1, \cdots, x_n)$. 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 $i, i = 1, \cdots, n$, maximizes his own payoff. Agent 0, however, does not only care about her own payoff, but also about the payoffs of all other n agents. Her objective function is $W(\mathbf{U})$, continuous and twice differentiable with $\mathbf{U} = (U_0, \cdots, U_n)$, $W_k \equiv \partial W/\partial U_k > 0$, $k = 0, \cdots, n$.

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

$$W_0 \frac{\partial U_0}{\partial y_i} + W_i \frac{\partial U_i}{\partial y_i} = 0 \tag{10}$$

Note that since $W_0, W_i > 0$ and $\partial U_i/\partial y_i > 0$, we must have $\partial U_0/\partial y_i < 0$ in the optimum. Differentiating $W(\mathbf{U})$ with respect to x_i, $i = 1, \cdots, n$, yields:

$$\sum_{k=0}^{n} W_k \frac{\partial U_k}{\partial x_i} = 0 \tag{11}$$

Whatever agent 0's precise preferences, her first best will always be on the payoff possibility frontier PPF. Every element $\mathbf{U}^*$ of the PPF is defined as:

$$U_i^*(\mathbf{U}_{-i}^*) \equiv \max_{\mathbf{x},\mathbf{y}} U_i \; s.t. \; \mathbf{U}_{-i} = \mathbf{U}_{-i}^*, \quad i = 1, \cdots, n \tag{12}$$

where $\mathbf{U}_{-i} \equiv (U_0, \cdots, U_{i-1}, U_{i+1}, \cdots, U_n)$. Let $\mathbf{x}^*$ be an $\mathbf{x}$ vector that is associated with a $\mathbf{U}^*$, and $\mathbf{X}^*$ the set of all $\mathbf{x}^*$:

$$\begin{aligned} \mathbf{x}^*(\mathbf{U}^*) &= \arg\max_{\mathbf{x}} U_i \; s.t. \; \mathbf{U}_{-i} = \mathbf{U}_{-i}^* \\ \mathbf{X}^* &\equiv \{\mathbf{x}^*(\mathbf{U}^*)\} \end{aligned} \tag{13}$$

In the following, we shall study the effect of sequential moves. The agents i, $i = 1, \cdots, n$, 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 interior solutions.[4] 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 shall assume that the second order conditions are satisfied.

In our interpretation of the Samaritan's dilemma and the rotten kid theorem, they do not only have 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. In subsection 3.4, we give the formal definitions and state the conditions for the Samaritan's dilemma to apply and for the rotten kid theorem to 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,

[4] This assumption is especially critical for Cornes and Silva's [12] public goods game, to be discussed in subsection 5.3. Chiappori and Werning [10] have shown that this game does not generally have an interior solution.

we shall derive the condition for the positive Samaritan's dilemma to hold:

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

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

$$\frac{\partial U_i}{\partial x_i} = 0 \tag{14}$$

In stage one, agent 0 sets the y_i that maximize her objective function $W(\mathbf{U})$, taking into account that the agents i's choices 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 (14) and differentiating it totally with respect to y_i, this can be rewritten as:

$$W_0 \frac{\partial U_0}{\partial y_i} + W_i \frac{\partial U_i}{\partial y_i} - \sum_{\substack{l=0 \\ l \neq i}}^{n} W_l \frac{\partial U_l}{\partial x_i} \frac{\partial^2 U_i / \partial y_i \partial x_i}{\partial^2 U_i / \partial x_i^2} = 0 \tag{15}$$

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.[5]

Condition 1 *For all $\mathbf{x} \in \mathbf{X}^*$, all $j = 1, \cdots, n$ and all $l = 0, \cdots, n$, $l \neq j$:*

$$\frac{\partial U_l}{\partial x_j} = 0$$

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. In general, this can only result in the first best for agent 0 if agent i's action does not affect

[5] All proofs are in the Appendix.

any other agent's payoff,[6] at least not on the margin. 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.

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

We solve the game by backwards induction, assuming an interior solution. In stage two, agent 0 sets the y_j that maximize her objective function $W(\mathbf{U})$, taking all x_i, $i = 1, \cdots, n$, as given:[7]

$$W_0 \frac{\partial U_0}{\partial y_j} + W_j \frac{\partial U_j}{\partial y_j} = 0 \tag{16}$$

In stage one, each agent i, $i = 1, \cdots, n$, sets the x_i that maximizes his own payoff, taking the x_l, $l = 1, \cdots, i-1, i+1, \cdots, n$, 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:

$$\frac{dU_i}{dx_i} \equiv \frac{\partial U_i}{\partial y_i} \frac{dy_i}{dx_i} + \frac{\partial U_i}{\partial x_i} = 0 \tag{17}$$

where the values for dy_j/dx_i, $j = 1, \cdots, n$, follow from the total differentiation of (16) with respect to x_i.

[6] This is the condition we already encountered in subsection 2.1.

[7] Obviously, these conditions are identical to the first order conditions (10) for agent 0's first best with respect to $\mathbf{y}$.

In general, the equilibrium condition (17) for $x_i,\ i = 1, \cdots, n$, is not identical to the corresponding first order condition (11) for agent 0's first best. We shall now see when it is. Let us first state an intermediate result.

Condition 2 *Consider a marginal change in x_i after which $\mathbf{y}$ is adjusted optimally according to (16). Define*

$$\frac{dU_0}{dx_i} \equiv \frac{\partial U_0}{\partial x_i} + \sum_{j=1}^{n} \frac{\partial U_0}{\partial y_j}\frac{dy_j}{dx_i} \tag{18}$$

$$\frac{dU_j}{dx_i} \equiv \frac{\partial U_j}{\partial x_i} + \frac{\partial U_j}{\partial y_j}\frac{dy_j}{dx_i} \tag{19}$$

for all $i, j = 1, \cdots, n$. Then $dU_l/dx_i = 0$ when $dU_i/dx_i = 0$ for all $\mathbf{x} \in \mathbf{X}^*$ and for all $l = 0, \cdots, n,\ i = 1, \cdots, n,\ l \neq i$.

Lemma 1 *Given that the second order conditions are satisfied, the positive rotten kid theorem holds for all $W(\mathbf{U})$ if and only if Condition 2 holds.*

Note the analogy with Proposition 1 from subsection 3.2. When agents i move last, they set $\partial U_i/\partial x_i = 0$. This will result in agent 0's first best for all $W(\mathbf{U})$ if and only if $\partial U_l/\partial x_i = 0$ for all other l. When agents i move first, they set $dU_i/dx_i = 0$. This will result in agent 0's first best for all $W(\mathbf{U})$ if and only if $dU_l/dx_i = 0$ for all other l.

Condition 2 is not a condition on the payoff functions $\mathbf{U}$ yet. We shall now derive such a condition.

Condition 3 *Take a vector $\mathbf{x}^*$ that implements a $\mathbf{U}^*$ according to (12). For this $\mathbf{x}^*$, write U_0 and U_i as:*

$$U_0(\mathbf{y}, \mathbf{x}^*) = G_0(\mathbf{x}^*) - F(\mathbf{z}) \tag{20}$$

$$U_i(y_i, \mathbf{x}^*) = G_i(\mathbf{x}^*) + z_i(y_i, \mathbf{x}^*) \tag{21}$$

with $\mathbf{z} \equiv (z_1, \cdots, z_n),\ \partial F/\partial z_i > 0,\ \partial z_i/\partial y_i > 0,\ i = 1, \cdots, n$. Then:

$$\sum_{l=1}^{n} \frac{\partial^2 F}{\partial z_j^* \partial z_l^*}\frac{\partial G_l}{\partial x_i} = 0 \tag{22}$$

Proposition 2 *Given that the second order conditions are satisfied, the positive rotten kid theorem holds for all $W(\mathbf{U})$ if and only if Condition 3 holds.*

In order to understand the implications of payoff condition 3, let us state:[8]

Lemma 2 1. *If and only if Condition 3 holds, there is a single vector $\mathbf{x}^*$ that implements the whole PPF.*

2. *This vector $\mathbf{x}^*$ maximizes aggregate income defined as:*

$$I(\mathbf{x}, \mathbf{y}) \equiv U_0(\mathbf{y}, \mathbf{x}) + \sum_{i=1}^{n} \frac{\partial U_0/\partial y_i}{\partial U_i/\partial y_i} U_i(y_i, \mathbf{x}) \tag{23}$$

for any $\mathbf{y}$ on the PPF.

We can regard

$$P_j \equiv -\frac{\partial U_0/\partial y_j}{\partial U_j/\partial y_j}$$

as the price of agent j's payoff to agent 0, because it denotes how much of U_0 agent 0 has to give up to increase U_j by one unit. Under Condition 3, the price of an agent j's payoff along the PPF is beyond manipulation by agent i: $dP_j/dx_i = 0$.[9]. Then we can aggregate all payoffs using these prices for a given $\mathbf{y}$ and refer to aggregate payoff as income $I(\mathbf{x}, \mathbf{y})$, as defined in (23). The agents i will maximize income and agent 0 will redistribute income. In the terminology of Monderer and Shapley [22], Condition 3 turns the game into a potential game, where all agents $i = 1, \cdots, n$ maximize the ordinal potential function $I(\mathbf{x}, \mathbf{y})$.

Let us define a Utility Possibility Curve UPC as the set of vectors $\mathbf{U}$ that can be obtained with a given $\mathbf{x}$. Then Lemma 2 implies that the whole Payoff Possibility Frontier PPF must consist of a single UPC. Figure 1, inspired by Bergstrom's [3] 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.

[8]The proof of Lemma 2.1 is contained in the proof of Proposition 2. With respect to Lemma 2.2, whereas the optimal value of $I(\mathbf{x}, \mathbf{y})$ depends on the choice of $\mathbf{y}$, the vector $\mathbf{x}^*$ does not depend on $\mathbf{y}$ under Condition 3.

[9]This is equation (27) from the proof of Proposition 2.

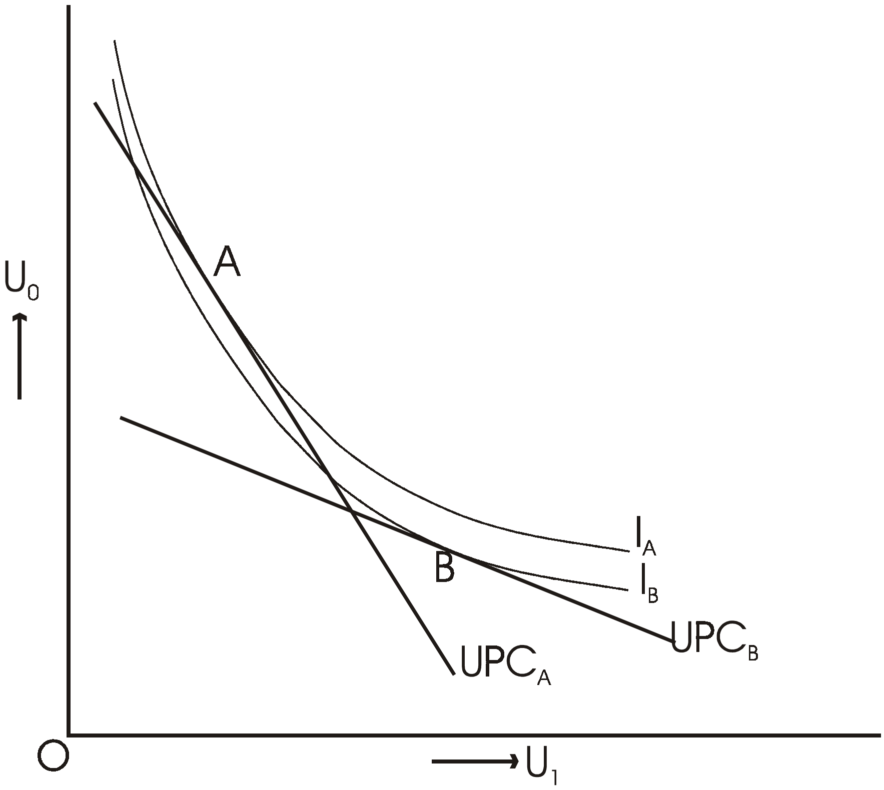

Figure 1: Intersecting Utility Possibility Curves

In Figure 1, point 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 agent 0's first-best indifference curve 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 will be higher than in point A. Thus, agent 1 will prefer implementing UPC_B to UPC_A.

In our introductory example of the rotten kid theorem (subsection 2.2), Condition 3 holds, because there is only one commodity, namely income.[10] Then we can identify payoffs with income and define aggregate or family income. All UPCs are parallel and

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

have slope -1, because they denote the feasible income distributions given aggregate 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. This means that the price of the parasite's payoff to the Samaritan is lower and the Samaritan will buy more of it.

3.4 Conditions for Samaritan's dilemma and rotten kid theorem

There are two alternative definitions for the Samaritan's dilemma and the rotten kid theorem, applied to agent 0's first best. The positive definition of the Samaritan's dilemma (the rotten kid theorem) is: agent 0 can reach her first best when she moves first (last). In subsections 3.2 and 3.3, we have already defined the positive versions of Samaritan's dilemma and rotten kid theorem and derived the conditions for them to hold.

The second definition of the Samaritan's dilemma (the rotten kid theorem) also includes a negative side: agent 0 can reach her first best when she moves first (last), but not when she moves last (first). This is the definition we adhere to in this paper. We shall now formally define the Samaritan's dilemma and the rotten kid theorem:

Definition 3 *The Samaritan's dilemma states that agent 0 can reach her first best when she moves in stage one and agents i, $i = 1, \cdots, n$, move in stage two, but not when agents i move in stage one and agent 0 moves in stage two.*

Definition 4 *The rotten kid theorem states that agent 0 can reach her first best when agents i move in stage one and agent 0 moves in stage two, but not when agent 0 moves in stage one and agents i, $i = 1, \cdots, n$, move in stage two.*

For the conditions under which the Samaritan's dilemma holds, we can simply take our Conditions 1 and 3:

Proposition 3 *Given that all agents' second order conditions are satisfied, the Samaritan's dilemma holds for all $W(\mathbf{U})$ if and only if Condition 1 holds and Condition 3 does not hold.*

For the rotten kid theorem, the analysis is somewhat more complicated. Agent 0 can reach her first best when she moves first for any $W_k > 0$ if all $\partial U_l/\partial x_j = 0,\ j = 1, \cdots, n,\ l = 0, \cdots, n,\ l \neq j$ (Condition 1). But substituting first best condition (10) into Condition 3, which ensures that agent 0 can reach her fist best when she moves last, reveals that the W_k are restricted to:

$$W_0 \frac{\partial F}{\partial z_i} = W_i$$

and first best condition (11) becomes:

$$\frac{\partial G_0}{\partial x_i} + \sum_{j=1}^{n} \frac{\partial F}{\partial z_j} \frac{\partial G_j}{\partial x_i} = 0$$

Then we don't need all individual $\partial U_l/\partial x_j = 0$, as long as the weighted sum is zero.[11]

Condition 4 *Given that Condition 3 holds:*

$$\frac{\partial G_0}{\partial x_i} - \frac{\partial F}{\partial z_i} \frac{\partial z_i}{\partial x_i} + \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{\partial F}{\partial z_j} \frac{\partial G_j}{\partial x_i} = 0$$

for $\mathbf{x} = \mathbf{x}^*$ and all $i = 1, \cdots, n$.

Proposition 4 *Given that all agents' second order conditions are satisfied, the rotten kid theorem holds for all $W(\mathbf{U})$ if and only if Condition 3 holds and Condition 4 does not hold.*

4 Pareto efficiency

Whereas we have stated the Samaritan's dilemma and the rotten kid theorem in terms of agent 0's first best, an alternative and often used definition is in terms of Pareto efficiency. In this section we shall examine the links between the two definitions.

[11] Obviously, Conditions 1 and 4 only differ for $n > 1$.

It is clear that agent 0's first best is a Pareto optimum, since any other allocation would make her worse off. The interesting question is: When agent 0 cannot reach her first best when she moves last (first), does this imply that the outcome is not Pareto efficient either?

Let us first establish the relation between agent 0's first best and Pareto efficiency in general. With equation (12) in subsection 3.1, we have already defined the payoff possibility frontier PPF. This definition implies:

Lemma 3 *For each allocation* $\mathbf{U}^*$ *on the Payoff Possibility Frontier (PPF),* $dU_k^*/dx_i = 0$ *is feasible for all* $k = 0, \cdots, n$*, and all* $i = 1, \cdots, n$*, where* dU_0/dx_i *and* dU_j/dx_i*,* $j = 1, \cdots, n$*, are defined by (18) and (19) respectively.*

The idea behind this lemma is the following. Consider a marginal change in x_i after which agent 0 adjusts $\mathbf{y}$ to compensate all agents j: $dU_j/dx_i = 0$ for all $j = 1, \cdots, n$. After this compensation, U_0 should also be back at its original level: $dU_0/dx_i = 0$. Otherwise, U_0 can be increased while all U_j, $j = 1, \cdots, n$, remain the same.

If agent 0 were selfish, then all allocations on the PPF would be Pareto efficient. However, when agent 0 is an altruist, a Pareto improvement from some allocations on the PPF may be possible. This would be the case if an increase in y_i, which obviously raises U_i, would also increase agent 0's objective function W. Let us now define the Altruistic Payoff Possibility Frontier $APPF$ as that part of the PPF from which Pareto improvements are impossible:

Definition 5 $\mathbf{U}^{A*}$ *is an element of agent 0's Altruistic Payoff Possibility Frontier (APPF) if and only if it is an element of the PPF and* $dW(\mathbf{U}^{A*})/dy_i \leq 0$ *for all* $i = 1, \cdots, n$*.*

Lemma 4 *All allocations and only the allocations* $\mathbf{U}^{A*}$ *are Pareto efficient.*

Let us now turn to the Samaritan's dilemma. In our introductory game in subsection 2.1, the sequence where the parasite moves first does not lead to a Pareto optimum. The parasite does not work hard enough, because a higher work effort would decrease the Samaritan's donation. Given the Samaritan's donation, however, the parasite could increase his own payoff and the Samaritan's objective function by working harder. We will now see that this result holds in general. Intuitively, the reason why a Pareto-efficient allocation is not agent 0's first best is that agent 0 does not like the payoff distribution. However, when agent 0 moves last, she determines the payoff distribution. Then, when the allocation is Pareto-efficient, it must be agent 0's first best.

Proposition 5 *In the game where agents $i, i = 1, \cdots, n$, move in stage one and agent 0 moves in stage two, the outcome is Pareto efficient if and only if it is agent 0's first best.*

In our continuous version of the rotten kid theorem (subsection 2.2), when the parent cannot reach her first best when she moves first, the outcome is not Pareto efficient either. Pareto efficiency requires the kid to maximize family income, but the kid will maximize his own income instead. We shall now see that this result can only be generalized partially.

Proposition 6 *Consider the game where agent 0 moves in stage one and agents $i, i = 1, \cdots, n$, move in stage two. The outcome is Pareto efficient if and only if it is agent 0's first best when:*

1. *$n = 1$, and/or:*
2. *agent 0 can reach her first best for all $W(\mathbf{U})$ when agents $i, i = 1, \cdots, n$, move in stage one and agent 0 moves in stage two.*

Otherwise, the resulting allocation $\mathbf{U}$ is a Pareto optimum if and only if it is on agent 0's $APPF$ according to Definition 5.

5 Bergstrom's rotten kid game

5.1 Introduction

The present paper is not the first to have derived conditions for the positive rotten kid theorem to hold. Bergstrom [3] and Cornes and Silva [12] have previously derived a condition from a more specific model than ours. In this 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 5.2, we shall introduce Bergstrom's [3] 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 an accordingly special version of our payoff condition. We shall also find that Bergstrom [3] was wrong in claiming that the payoff condition is necessary only when "money is important enough". In subsection 5.3, we shall discuss Cornes and Silva's [12] condition for the positive rotten kid theorem to hold in Bergstrom's [3] 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 [3] model.

5.2 Bergstrom's solution

In this subsection, we shall discuss Bergstrom's [3] conditions for the positive rotten kid theorem. In Bergstrom's [3] 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 [3]. 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 6 *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* $U_0(\mathbf{y}, \mathbf{x})$ *is given by* $U_0(y_0, \mathbf{x})$ *with* $y_0 \equiv -\sum_{i=1}^n y_i$.

Let us now apply this Definition to our Condition 3 for the positive rotten kid theorem. For the payoff functions (20) and (21), it implies that $z_i(y_i, \mathbf{x}) = A(\mathbf{x})y_i$ for all $i = 1, \cdots, n$ and $F(\mathbf{z}) = \sum_{i=1}^n z_i$. Thus all $\partial^2 F/\partial z_i \partial z_l$ are equal. Then there are two ways to satisfy condition (22). The first is to set the $\partial^2 F/\partial z_i \partial z_l = 0$. Then Condition 3 becomes:

$$U_k = A(\mathbf{x})y_k + B_k(\mathbf{x}) \tag{24}$$

for all $k = 0, \cdots, n$, with $y_0 \equiv -\sum_{i=1}^n y_i$.

The second way to satisfy (22) is to set all $\partial G_l/\partial x_i = 0$. Then $\partial G_0/\partial x_i = 0$ for all $i = 1, \cdots, n$ in agent 0's first best given by (10) and (11). Since there is only one $\mathbf{x}^*$ that implements the whole PPF, this means that the function $G_0(\mathbf{x})$ is a constant on the PPF. Then we can apply a monotonous transformation to U_0 and write it as $U_0 = -\sum_{i=1}^{n} z_i = A(\mathbf{x})y_0$. This is just a special form of (24) for $k = 0$. Then all $U_i, i = 1, \cdots, n$, must have the form (21). Thus, given that $\mathbf{y}$ is money, the positive rotten kid theorem holds for all W if and only if all $U_k,\ k = 0, \cdots, n$ satisfy (24).

The second step from our framework to Bergstrom's [3] is that the budget constraint on y_0 is binding. The third and final step is to exclude U_0 from agent 0's objective function. After these three stepts, agent 0's maximization problem is:

$$\max\ W(U_1(y_1, \mathbf{x}), \cdots, U_n(y_n, \mathbf{x})) \qquad s.t.\ \sum_{i=1}^{n} y_i = y(\mathbf{x}) \tag{25}$$

The second and third steps do not result in a further restriction on the payoff functions. Thus, in the game (25), the positive rotten kid theorem holds for all $W(\mathbf{U})$ and all $y(\mathbf{x})$ if and only if U_k has the form (24) for $k = 1, \cdots, n$. This is exactly the condition that Bergstrom [3] derives. Our condition 3 is more general than Bergstrom's [3], but that is only because we haven't restricted agent 0's actions to giving money.

How does restricting $\mathbf{y}$ to money result in a further restriction on the payoff functions? As we know from susbsection 3.3, 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 only depends 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 $\mathbf{x}$. Or, 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.

We have now shown that condition (24) is necessary and sufficient for the positive rotten kid theorem to hold. Bergstrom [3], 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 3.2. 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.[12] 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 [3] condition that money is important enough can be stated as follows:

Condition 5 *$\partial U_i(y_i, \mathbf{x}) > 0$ for all $y_i > 0$, $i = 1, \cdots, n$, 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 $U_i(y_i', \mathbf{x}^0) = U_i(y_i, \mathbf{x})$.*

Bergstrom [3] 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 $i, j = 1, \cdots, n$. However, all that is needed to prove this is the first part of Condition 5 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 only depends 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.

5.3 Cornes and Silva's solution

In subsection 5.2, we have seen that Bergstrom's [3] own condition (24) for the positive rotten kid theorem in his game is a special case of our Condition 3. Cornes and Silva [12] recently found another and completely different condition for the positive rotten kid theorem to hold in Bergstrom's [3] framework. Under this condition, all kids contribute to a pure public good. In this subsection, we shall argue that the reason why Cornes and Silva [12] could find an additional condition is that they, unlike Bergstrom [3], have put a restriction on the budget constraint.

[12] Obviously, if one does not take the second order conditions for granted, payoff condition (24) is not sufficient for the positive rotten kid theorem to hold. Bergstrom [3] overlooks this point in his Proposition 1 (p. 1148) which effectively states that the positive rotten kid theorem holds when payoffs satisfy (24).

We shall first discuss Cornes and Silva's [12] 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 [3] framework, nor in our more general setup.

In the notation of this paper, Cornes and Silva's [12] model can be described as follows. Agent i, $i = 1, \cdots, n$, only affects the others through his contribution x_i to a pure public good $X \equiv \sum_{i=1}^{n} x_i$. Agent i has to decide how much x_i of his initial exogenous endowment m_i to contribute to the pure public good. The rest of the endowment plus the transfer t_i from agent 0 is available for consumption y_i of the private good. Agent 0's budget is zero: $\sum_{i=1}^{n} t_i = 0$, so that she will also take away from some agents: $t_i < 0$ is feasible. 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 [12] manage to find this additional solution? To find that out, let us first briefly present the derivation of Bergstrom's [3] own solution with our method from subsection 3.3. Analogous to Lemma 1, $dU_j/dx_i = 0$ must hold for all $i, j = 1, \cdots, n$ 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 $l, i = 1, \cdots, n$, $l \neq i$. These conditions are (24).

How can we possibly find an additional payoff condition for all $W(\mathbf{U})$? Obviously, this condition should also yield $dU_l/dx_i = 0$ for all $l, i = 1, \cdots, n$, $l \neq i$. In deriving condition (24), we have assumed that agent 0 would have to set all $dU_l/dx_i = 0$ herself. Alternatively, we could impose some restrictions R on the payoff functions $U_i(y_i, \mathbf{x})$ so that $dU_i/dx_i = 0$ automatically implies $dU_l/dx_i = 0$ for some (but not all) $l, i = 1, \cdots, n$, $l \neq i$. 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 $dU_i/dx_l = 0$ for all $l, i = 1, \cdots, n$, $l \neq i$. 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 $dU_i/dx_j = 0$, $i, j = 1, \cdots, n$, 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 [12] pure public good case. Note that this condition does not only put restrictions on the agents' payoff functions $U_i(y_i, \mathbf{x})$, but also on the budget constraint $y = y(\mathbf{x})$. Bergstrom's [3] own condition is necessary and sufficient for the positive rotten kid theorem to hold in the game (25) for all $W(\mathbf{U})$ and all $y(\mathbf{x})$. If we allow for restrictions on $y(\mathbf{x})$, then there is exactly one additional condition, which is Cornes and Silva's [12].

Following the above reasoning, it is clear why Cornes and Silva's [12] condition does not carry over to our more general framework. In the pure public good case where $X \equiv \sum_{i=1}^n x_i$, the agents i, $i = 1, \cdots, n$, will set $dU_i/dX = 0$. However, this is not sufficient. 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 3 with $\mathbf{x}$ replaced by $X \equiv \sum_{i=1}^n x_i$. Thus, it is impossible to find any solution other than Condition 3 in the general framework.

6 Conclusion

For twenty-five years, the Samaritan's dilemma (Buchanan [7]) and the rotten kid theorem (Becker [1] [2]), 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 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 payoffs.

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.[13] It would be worthwhile to expand the general analysis to encompass multi-period models and incomplete or asymmetric in-

[13] Bruce and Waldman [5] [6] and Lindbeck and Weibull [20] have analyzed two-period lifetime models. Chami [8] [9] and Lagerlöf [19] assume asymmetric information. Coate [11], Lord and Raganzas [21] and Wigger [27] include uncertainty.

formation.

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 some other player. Thus, the theory of altruism can contribute to our understanding of when collective and individual interests coincide (Shapiro and Petchey [26], Munger [23]). 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, the time-consistent policy even results in the first best. Starting with Kydland and Prescott [18], most analyses of time consistency have used a more complicated setup than ours.[14] However, a general framework for the study of time consistency issues is still lacking. Our simple model of altruism would be a useful starting point for the development of such a framework (Dijkstra [15]).

7 Appendix

Proof of Proposition 1. Combining Condition 1 with agents i's first order conditions for the maximization of U_i (14), we obtain the first best conditions for $\mathbf{x}$ (11). Substituting Condition 1 into agent 0's first order conditions for the maximization of W (15), we obtain the first best conditions for $\mathbf{y}$ (10). This proves the "if" part. The "only if" part follows from the requirement that (14) and (15) should turn into (11) and (10) for all values of $W_k > 0$. This is only possible when Condition 1 holds.

Proof of Lemma 1. 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) into the first best conditions (11) for $\mathbf{x}$, we can rewrite them as:

$$\sum_{k=0}^{n} W_k \frac{dU_k}{dx_i} = 0 \tag{26}$$

for all $i = 1, \cdots, n$, where dU_k/dx_i, $k = 0, \cdots, n$, is defined by (18) and (19).

[14] Dijkstra [14] offers a straightforward application of the Samaritan's dilemma to time consistency.

In the equilibrium of the game, agent i, $i = 1, \cdots, n$, sets $dU_i/dx_i = 0$. This will result in the first best condition (26) for all $W_k > 0$, $k = 0, \cdots, n$, if and only if $dU_i/dx_i = 0$ implies $dU_l/dx_i = 0$ for all $i = 1, \cdots, n$, $l = 0, \cdots, n, l \neq i$ in agent 0's first best, characterized by $\mathbf{x} \in \mathbf{X}^*$. This is Condition 2.

Proof of Proposition 2. By Lemma 1, $dU_k/dx_i = 0$, $k = 0, \cdots, n$, $i = 1, \cdots, n$, must hold in agent 0's first best. To find the expressions for dU_k/dx_i, write the total differential of agent 0's first order condition (16) with respect to x_i:

$$\frac{\partial U_0}{\partial y_j}\sum_{k=0}^{n} W_{0k}\frac{dU_k}{dx_i} + W_0\frac{d(\partial U_0/\partial y_j)}{dx_i} + \frac{\partial U_j}{\partial y_j}\sum_{k=0}^{n} W_{jk}\frac{dU_k}{dx_i} + W_j\frac{d(\partial U_j/\partial y_j)}{dx_i} = 0$$

Substituting $dU_k/dx_i = 0$, $k = 0, \cdots, n$, $i = 1, \cdots, n$, this reduces to:

$$W_0\frac{d(\partial U_0/\partial y_j)}{dx_i} + W_j\frac{d(\partial U_j/\partial y_j)}{dx_i} = 0$$

Substituting agent 0's first order conditions (16) for $\mathbf{y}$, this becomes:

$$d\left(\frac{\partial U_0/\partial y_j}{\partial U_j/\partial y_j}\right)/dx_i = 0 \tag{27}$$

This means that the slope of the Payoff Possibility Frontier (PPF) does not change with a change in $\mathbf{x}$. Thus, there is only one $\mathbf{x}$, which we call $\mathbf{x}^*$, that implements the whole PPF (*Lemma 2*).

Given $\mathbf{x} = \mathbf{x}^*$, we can write the agents' payoffs as (21) and (20). Replacing y_j by z_j and substituting (21) and (20), condition (27) becomes condition (22):

$$\frac{d\left(\partial F/\partial z_j\right)}{dx_i} = \sum_{l=1}^{n}\frac{\partial^2 F}{\partial z_j \partial z_l}\frac{dz_l}{dx_i} = -\sum_{l=1}^{n}\frac{\partial^2 F}{\partial z_j \partial z_l}\frac{\partial G_l}{\partial x_i} = 0$$

The second equality follows from the fact that $dU_l/dx_i = 0$ for all $i, l = 1, \cdots, n$ by Lemma 1.

Proof of Proposition 5. The "if" part is obvious. With respect to the "only if" part, note that agent 0 maximizes W with respect to $\mathbf{y}$ according to (10) in stage two. By Lemmas 3 and 4, there exist dy_j/dx_i for all $i = 1, \cdots, n$ such that a Pareto optimum satisfies:

$$W_0\left(\frac{\partial U_0}{\partial x_i} + \sum_{j=1}^{n}\frac{\partial U_0}{\partial y_j}\frac{dy_j}{dx_i}\right) + \sum_{j=1}^{n} W_j\left(\frac{\partial U_j}{\partial x_i} + \frac{\partial U_j}{\partial y_j}\frac{dy_j}{dx_i}\right) = 0$$

Substituting (10), this becomes:

$$\sum_{k=0}^{n} W_k \frac{\partial U_k}{\partial x_i} = 0$$

These are (11), the first order conditions for W with respect to $\mathbf{x}$. Thus, all first best conditions are satisfied and the allocation is agent 0's first best.

Proof of Proposition 6. In Case 1, by Lemmas 3 and 4, $\mathbf{U}$ can only be Pareto efficient if $dU_0/dx = dU_1/dx = 0$ is feasible. For dU_1/dx, we find:

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

The second equality follows from the fact that agent 1 sets $\partial U_1/\partial x = 0$ in stage two. Thus, $dU_1/dx = 0$ implies $dy/dx = 0$. Then for dU_0/dx:

$$\frac{dU_0}{dx} = \frac{\partial U_0}{\partial x} + \frac{\partial U_0}{\partial y}\frac{dy}{dx} = \frac{\partial U_0}{\partial x}$$

Thus, $dU_0/dx = dU_1/dx = 0$ is feasible if and only if $\partial U_0/\partial x = 0$. But then Condition 1 is satisfied and $\mathbf{U}$ is agent 0's first best.

In Case 2, Lemma 2 implies that the $\mathbf{x}^*$ that implements agent 0's first best implements the whole PPF. When agent 0 cannot reach her first best when she moves first, $\mathbf{x} \neq \mathbf{x}^*$ and the allocation is not on the PPF. By Lemma 4, when an allocation is not on the PPF, it is not Pareto efficient either.

References

[1] Becker, Gary S. (1974), "A theory of social interaction", Journal of Political Economy 82: 1063-1093.

[2] Becker, Gary S. (1976), "Altruism, egoism, and genetic fitness: Economics and sociobiology", Journal of Economic Literature 14: 817-826.

[3] Bergstrom, Theodore C. (1989), "A fresh look at the rotten kid theorem— and other household mysteries", Journal of Political Economy 97: 1138-1159.

[4] Bernheim, B. Douglas, Andrei Schleifer and Lawrence H. Summers (1985), "The strategic bequest motive", Journal of Political Economy 93: 1045-1076.

[5] Bruce, Neil and Michael Waldman (1990), "The rotten-kid theorem meets the Samaritan's dilemma", Quarterly Journal of Economics 105: 155-165.

[6] Bruce, Neil and Michael Waldman (1991), "Transfers in kind: Why they can be efficient and nonpaternalistic", American Economic Review 81: 1345-1351.

[7] Buchanan, James M. (1975), "The Samaritan's dilemma", in: E.S. Phelps (ed.), Altruism, Morality, and Economic Theory, Sage Foundation, New York, 71-85.

[8] Chami, Ralph (1996), "King Lear's dilemma: Precommitment versus the last word", Economics Letters 52: 171-176.

[9] Chami, Ralph (1998), "Private income transfers and market incentives", Economica 65: 557-580.

[10] Chiappori, Pierre-André and Iván Werning (2002), "Comment on "Rotten kids, purity, and perfection"", Journal of Political Economy 110: 475-480.

[11] Coate, Stephen (1995), "Altruism, the Samaritan's dilemma, and government transfer policy", American Economic Review 85: 46-57.

[12] Cornes, Richard C. and Emilson C.D. Silva (1999), "Rotten kids, purity, and perfection", Journal of Political Economy 107: 1034-1040.

[13] Cox, Donald (1987), "Motives for private income transfers", Journal of Political Economy 95: 508-546.

[14] Dijkstra, Bouwe R. (2000), "Investment incentives of environmental policy instruments", Discussion Paper 308, Faculty of Economics, University of Heidelberg.

[15] Dijkstra, Bouwe R. (2002), "Policy commitment vs. time consistency", mimeo, School of Economics, University of Nottingham.

[16] Hirshleifer, Jack (1977), "Shakespeare vs. Becker on altruism: The importance of having the last word", Journal of Economic Literature 15: 500-502.

[17] Jürges, Hendrik (2000), "Of rotten kids and Rawlsian parents: The optimal timing of intergenerational transfers", Journal of Population Economics 13: 147-157.

[18] Kydland, Finn E. and Edward C. Prescott (1977), "Rules rather than discretion: The inconsistency of optimal plans", Journal of Political Economy 85: 473-491.

[19] Lagerlöf, Johan (1999), "Incomplete information in the Samaritan's dilemma: The dilemma (almost) vanishes", Discussion Paper FS IV 99-12, Wissenschaftszentrum Berlin.

[20] Lindbeck, Assar and Jörgen W. Weibull (1988), "Altruism and time consistency: The economics of fait accompli", Journal of Political Economy 96: 1165-1182.

[21] Lord, William and Peter Raganzas (1995), "Uncertainty, altruism, and savings: Precautionary savings meets the Samaritan's dilemma", Public Finance 50: 404-419.

[22] Monderer, Dov and Lloyd S. Shapley (1996), "Potential games", Games and Economic Behavior 14: 124-143.

[23] Munger, Michael C. (2000), "Five questions: An integrated research agenda for Public Choice", Public Choice 103: 1-12.

[24] Pollak, Robert A. (1985), "A transactions cost approach to families and households", Journal of Economic Literature 23: 581-608.

[25] Schmidtchen, Dieter (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.

[26] Shapiro, Perry and Jeffrey Petchey (1998), "The coincidence of collective and individual interests", Working paper 9-98R, Department of Economics, University of California, Santa Barbara.

[27] Wigger, Berthold U. (1996), "Two-sided altruism, the Samaritan's dilemma, and universal compulsory insurance", Public Finance 51: 275-290.