Altruism in the Principal-Agent Model: The Samaritan's Dilemma Revisited*

Suman Ghosh and Alexander Karaivanov†

March, 2008

Abstract

We analyze the Samaritan's Dilemma problem in a principal-agent setting in which transfers from an altruist to the agent have an effect on the agent's contractual relationship with the principal. We show that the presence of the altruist can affect the terms of the principal-agent contract in a way so that the standard inefficiency result of effort underprovision in the Samaritan's Dilemma context is completely avoided or at least mitigated. The result remains robust if the transfer from the altruist is endogenized.

Keywords: the Samaritan's Dilemma; moral hazard; altruism

JEL Classifications: D64, D82

*We thank Dilip Mookherjee, Eric Van Tassel, and Mike Waldman for their extremely helpful comments and suggestions. All remaining errors and omissions are ours.

†Ghosh: Department of Economics, Florida Atlantic University, 777 Glades Road, Boca Raton, FL 33431, USA; email: sghosh@fau.edu. Karaivanov: Department of Economics, Simon Fraser University, 8888 University Drive, Burnaby, BC, V5A 1S6, Canada; email: akaraiva@sfu.ca.

1 Introduction

The Samaritan's Dilemma[1] states that if an altruist promises to help out an agent in the event of a bad outcome, this promise will negatively affect the agent's effort or the savings that the agent might otherwise accumulate. Thus, the altruist faces a dilemma: on the one hand, he wants to help the agent in bad times, while, on the other hand, he must acknowledge the detrimental effect of his help on the agent's incentives and ex-post outcomes.

The importance of the Samaritan's Dilemma effect has been widely recognized and studied in various contexts. For instance, the inefficiency result has been used by Kotlikoff (1987), Lindbeck and Weibull (1988) and Hansson and Stuart (1989) to justify the existence of compulsory social insurance systems. The basic argument is that, assuming the presence of a Samaritan's Dilemma problem, these compulsory schemes effectively force people to save more than what they would have done voluntarily. Also, in an infinite horizon overlapping generations model, O'Connell and Zeldes (1993) argue that the strategic under-saving result present in the Samaritan's Dilemma problem, has consequences for the overall dynamic efficiency of the economy since less savings lead to lower capital stock and hence affect the interest rate. In yet another vein, Bruce and Waldman (1990, 1991) and Coate (1995) build on the basic idea of Samaritan's Dilemma and show that in such setting in-kind or tied transfers might be efficiency enhancing. Their reasoning is that by making the aid transfer an illiquid investment, (e.g. parents providing college education for the child) as opposed to an equivalent cash payment, the altruist can ensure that the negative incentive effect does not arise. Finally, Lagerlof (2004) shows that the under-saving result is mitigated if one relaxes the assumption of complete information. The intuition is that if the altruist is uncertain about the magnitude of the agent's need for aid, the agent has an incentive to signal that he is in great need by saving more than he would otherwise do.

[1] See Buchanan (1975) for an early formal treatment.

The above-mentioned papers study the Samaritan's Dilemma problem in the context of the interaction between an altruist and an agent only. However, in many applications there is need for an expanded framework of the altruist-agent (e.g. parent-child) interaction. In their extensive literature survey, Haveman and Wolfe (1995) argue that too much emphasis has been placed on the parent's (altruist's) choices, while not enough attention is paid to the child's (agent's) own choices as well as to society's (or, in general, third parties') role, including market forces. In our view, a complete model of the Samaritan's Dilemma problem must integrate third parties into the altruist-agent relationship, each with their utility maximizing exercise. This paper proposes such a theoretical framework. Specifically, we analyze situations in which the transfer from the altruist has a bearing on the agent's contractual relationship with a third party (a principal). We show that the terms of the principal-agent contract are affected by the actions of the altruist and can change in a way so that the inefficiency result in the typical Samaritan's dilemma is mitigated or even completely avoided.

To motivate our story further, consider two specific examples of principal-agent settings where Samaritan's Dilemma type of problems arise. First, in the development economics literature on credit market imperfections Ghatak, Morelli and Sjostrom (2001) show that, in an overlapping generations model with credit constraints, agents who work hard and save enjoy later advantages over agents who rely on external finance. The authors therefore argue that public policies aimed at reducing credit market imperfections can negatively influence the dynamic incentives for young workers to work hard and save. This scenario maps well to our setting as our altruist can be interpreted as the government or a microfinance institution whose objective is providing a safety net for unsuccessful agents. Second, our results are in line with the findings of Ghosh, Mookherjee and Ray (2001) who point out that policies like the altruist's transfers in our model which empower agents (borrowers in their model) through strengthening their bargaining position with the principal (the lender in their model) are likely to increase the agents' effort level and efficiency.

Our theoretical setting is the standard principal-agent model. The agent has a production technology with stochastic output (success or failure), the probability of which is a function of the agent's effort level. Agent's effort is unobservable to the principal which results in a moral hazard problem. Output is observable. The principal pays the agent a wage conditional on output. Additionally, there is an enforcement (commitment) problem so that, upon success, the principal can obtain the output from the agent only if he pays a wage at least equal to the agent's outside option of consuming/selling the output[2].

We first demonstrate the standard Samaritan's Dilemma outcome whereby the introduction of an altruist in the principal-agent model who gives handouts to failed agents reduces the agents' effort level. We then analyze how the presence of the altruist makes the principal change the optimal wage contract offered to the agent. We consider two cases: one in which the altruist observes output and one in which he does not. In the former case, the agent's effort unambiguously decreases. In the latter case, agents (whether successful or failed) can always claim failure and go to the altruist for the handout. Because of this, the principal must ensure that the agent's participation constraint takes into account this possibility. We show that the payment which the agent receives from the principal in the event of success increases so that the negative effect on the effort level due to the presence of the Samaritan is mitigated if the agent is risk averse or completely offset if the agent is risk neutral.

The paper is organized as follows. Section 2 describes the basic principal-agent model. We then introduce the altruist and analyze how his presence affects the optimal principal-agent, agent's effort, and efficiency. An extension to the basic model in which we endogenize the altruist's transfer decision is presented in section 4. The last section concludes.

[2]This assumption simplifies the analysis because we need to deal with ex-post rather than ex-ante participation constraints for the agent. The main mechanism behind our results is, however, not affected.

2 Model

Consider an economy consisting of a principal and an agent. The agent supplies effort, e towards a productive task which yields positive output (success) with probability $p(e)$ and zero output (failure) with probability $1 - p(e)$. The function $p(e)$ is increasing and strictly concave with $p(0) = 0$. The agent's utility is $u(c) - g(e)$, where u is increasing and (weakly) concave and g is increasing and convex with $g(0) = 0$. The output has value V to the principal but only a value of $X < V$ to the agent. The agent's effort, e is not observable by the principal while output is perfectly observable. In addition, we assume limited commitment, i.e. the agent cannot commit ex-ante not to hold up the principal and require a payment of at least X in order to release the output if production is successful (see footnote 2). In other words, the agent can, after observing a successful outcome, consume the output himself. Thus X can be thought of as the reservation utility of the agent, given that he is successful. If the agent is unsuccessful it is assumed that his (ex-post) outside option is zero.

Because agent's effort is unobservable, a moral hazard problem arises in the principal-agent relationship. The principal has to decide what wage to pay the agent in case of success and in case of failure, w_H and w_L respectively. To make the moral hazard problem relevant (even if the agent is risk-neutral), assume limited liability, i.e. $w_L, w_H \geq 0$ – punishing the agent with a negative wage is not possible. Finally, assume that the agent cannot produce anything on his own, i.e. his ex-ante reservation utility is zero.

2.1 The Agent's Problem

Given the wages w_H, w_L the agent solves:

$$\max_{e} \; p(e)u(w_H) + (1 - p(e))u(w_L) - g(e)$$

The first order condition for an interior solution is:

$$p'(e)[u(w_H) - u(w_L)] = g'(e) \tag{1}$$

The optimal effort level, e^* depends on the utility differential induced by both wages. Specifically, note that if $w_H = w_L$ we have $e^* = 0$. That is, because of the moral hazard problem, the agent will not supply effort unless he is given the right incentives. As is standard in such settings, the agent needs to be rewarded for high output and punished for low output. Given our assumptions on g and p, we have that $F(e) \equiv \frac{g'(e)}{p'(e)}$ is an increasing function in e and hence the optimal effort level is increasing in the utility differential.

2.2 The Principal's Problem

The principal chooses w_H and w_L optimally and recommends an effort level e, given the agent's incentive compatibility constraint. Since there are only two possible output levels and $p(e)$ is strictly concave, the first-order approach (Rogerson, 1985) is valid[3], so we can use the agent's first order condition (1) as the principal's incentive compatibility constraint (ICC).

Given the commitment problem, in addition to the ICC, the principal needs to take into account the agent's (ex-post) participation constraints (PC) for each state. That is, he will need to provide the agent with at least $u(X)$ in case of success and $u(0)$ in case of failure[4]. The principal's problem is therefore:

$$\max_{w_H, w_L, e} \; p(e)(V - w_H) - (1 - p(e))w_L$$

subject to:

$$p'(e)[u(w_H) - u(w_L)] = g'(e) \tag{2}$$

$$u(w_H) \geq u(X) \tag{3}$$

$$u(w_L) \geq u(0) = 0 \tag{4}$$

Note first that the participation constraint for the low output state, (4) must always

[3]See Karaivanov (2007) for a proof.

[4]Assume $u(0) = 0$ for simplicity and to make effort interior.

bind in the optimal contract since, ceteris paribus, increasing w_L reduces the profits for the principal both directly (he needs to pay the agent more) and also indirectly (by reducing agent's effort, and hence expected output). This implies that $w_L = 0$ – a failed agent is given the maximum feasible punishment under limited liability. The limited commitment problem rules out the possibility of the principal insuring the risk-averse agent (i.e. paying $w_L > 0$ and $w_H < X$). Second, because of the moral hazard problem, the ICC (2) must always bind for a positive effort level – incentive-compatibility needs to be ensured.

However, there are two possibilities regarding the participation constraint for high output, (3) – it can either bind, and so $w_H = X$ or not bind, i.e. $w_H > X$. In the latter case, the intuition is that increasing the reward for success w_H raises e and so, if X is low enough, setting $w_H > X$ might generate higher profits than giving the minimum required pay of X to the agent.

Let ϕ_1 be the function mapping w_H into e from (2) i.e. the solution to $u(w_H) = g'(e)/p'(e)$. We assume that

Assumption A1: $p(\phi_1(w_H))$ *is concave in* w_H.

Basically, we require the probability of success at the equilibrium effort level, $p(e^*)$ to be increasing but at a decreasing rate in the wage w_H – i.e. "decreasing returns" in the reward. The following proposition characterizes the optimal wage contract.

Proposition 1:

(a) If

$$p'(\phi_1(X))\phi_1'(X)(V - X) - p(\phi_1(X)) \leq 0 \tag{5}$$

then (3) binds, the optimal wages are $w_L^* = 0$, $w_H^* = X$, *and the optimal effort* *level,* $e^* = \phi_1(X)$ *solves*

$$u(X) = \frac{g'(e)}{p'(e)} \tag{6}$$

(b) If, instead, $p'(\phi_1(X))\phi_1'(X)(V - X) - p(\phi_1(X)) > 0$, *then the optimal wages are* $w_L^* = 0$, $w_H^* = \tilde{w}^* > X$ *and* $e^* = \phi_1(\tilde{w}^*)$.

Proof: (see Appendix)

The inequality in Proposition 1(a) states that the derivative of the principal's profit function is non-positive at $w_H = X$. That is, the optimal wage in case of success is lower than X, the PC binds, and hence we have a corner solution. The opposite is true in part (b) when the PC does not bind and the solution for w_H is interior.

3 The Principal-Agent Problem in the Presence of an Altruist

We now introduce a Samaritan (altruist) in the model. The altruist derives utility from bailing out the agent if the latter fails. We study how the introduction of the altruist affects the optimal principal-agent wage contract, as well as the associated effort level supplied by the agent. We first assume that the altruist observes the outcome of the project and hence can condition his aid transfer accordingly. We call this case a "safety net" altruist. We then study the case in which the altruist cannot observe the outcome of the project (a "non-discriminating altruist").[5]

3.1 A "Safety Net" Altruist

In this case the altruist observes the agent's output and gives him an amount $Y > 0$ upon failure. No help is provided upon success. The principal's problem is modified by the different

[5] As we will see later, depending on his objectives, the altruist could knowingly ignore the outcome of the project and provide the transfer in both states. Interestingly, we show that this might be efficiency enhancing.

ICC and also the modified participation constraint upon failure which now has to take into account the transfer from the altruist:

$$u(w_L) \geq u(Y)$$

The agent's problem in the presence of the altruist thus becomes:

$$\max_e \; p(e)u(w_H) + (1 - p(e))u(w_L) - g(e)$$

with a FOC:

$$p'(e)[u(w_H) - u(w_L)] = g'(e) \tag{7}$$

As before, it is optimal for the principal to set $w_L = Y$ (the lowest feasible value). Also, the ICC has to bind again but the participation constraint in the high state might not. Define ϕ_2 as the function mapping w_H into e (given Y) from the equation:

$$u(w_H) - u(Y) = \frac{g'(e)}{p'(e)} \tag{8}$$

Proceeding as in the proof of Proposition 1, whether (3) binds or not, depends on the sign of the derivative with respect to w_H of principal's profits evaluated at $w_H = X$:

$$p'(\phi_2(X))\phi_2'(X)(V - X + Y) - p(\phi_2(X)) \tag{9}$$

If the expression in (9) is non-positive, we have a corner solution, $w_H = X$; otherwise we have an interior solution ($w_H > X$). Denote the solution to the above problem by $e^{**} \equiv \phi_2(w_H)$. Comparing the optimal contracts in the cases with and without the altruist gives:

Proposition 2: *The agent's effort level in the presence of the altruist, e^{**} is strictly smaller than the agent's effort level in the original principal-agent problem without the altruist, e^*.*

Proof: Denote by w_H^{**} the optimal high-output wage in the presence of the altruist. Note first that the condition for (9) to be ≤ 0 is stronger than the corresponding condition in (5) because, for any w, we have $\phi_2(w) < \phi_1(w)$ and hence $p'(\phi_2(w)) > p'(\phi_1(w))$, while $\phi_1'(w) = \phi_2'(w)$ and $Y > 0$. Thus, a higher wage is needed to make the derivative in (9) negative. Therefore, (9) being negative implies (5) being negative. Given this result, there are two possible cases:

(1) If $p'(\phi_2(X))\phi_2'(X)(V - X + Y) - p(\phi_2(X)) \leq 0$ then $w_H^{**} = w_H^* = X$ and so the effort level e^{**} is less than e^* because at e^* the LHS of (8) evaluated at $w_H = X$ is smaller than the RHS (which equals $u(X)$ from (6)), and so e needs adjust downwards to decrease the RHS and restore optimality (remember $\frac{g'(e)}{p'(e)}$ is increasing in e).

(2) If $p'(\phi_2(X))\phi_2'(X)(V-X+Y)-p(\phi_2(X)) > 0$, then Proposition 1 implies that the PC in the high state does not bind, so w_H^{**} equals some value, $\tilde{w}^{**}$ such that $\tilde{w}^{**} > w_H^* \geq X$. Let $\hat{w}$ be the wage level such that $u(w_H^{**}) - u(Y) = u(\hat{w})$ and let as before $\tilde{w}^*$ solve $p'(\phi_1(w))\phi_1'(w)(V - w) - p(\phi_1(w)) = 0$ (that is, $\tilde{w}^*$ is the profit maximizing wage when there is no altruist and the agent's PC is not binding). In the presence of the altruist, at the optimal effort level e^{**} (provided through w_H^{**} and Y), we have that the marginal benefit for the principal of raising w_H by a small amount is $MB^{**} \equiv p'(e^{**})\frac{\partial e}{\partial w_H}(V - w_H^{**} + Y)$ which, given the interior equilibrium, equals the marginal cost of raising the wage, $MC^{**} \equiv p(e^{**})$.

Now go back to the case without an altruist. The same effort level e^{**} as above can be provided by setting $w^H = \hat{w}$ and $w^L = 0$. But then we have that the marginal benefit of raising w_H at that effort level exceeds the marginal cost of doing so, i.e. $MB^* > MC^* = MC^{**}$. The latter holds since $MB^* > MB^{**}$ which is true because $\frac{\partial e}{\partial w_H}$ is larger evaluated at $\hat{w}$ vs. at $\tilde{w}^{**}$ (compare (6) and (8)), and because $w_H^{**} - Y \geq \hat{w}$. Both these statements follow by the concavity of u. Thus, if the PC in the case without the altruist does not bind (an interior solution), it must be that $w_H^* = \tilde{w}^* > \hat{w}$ (to reduce the marginal benefit and raise the marginal cost of effort, the wage $\tilde{w}^*$ must be higher than $\hat{w}$) and hence $e^* > e^{**}$. If the PC binds, then $X = w_H^* > \tilde{w}^*$, and so again $e^* = \phi_1(X) > \phi_1(\tilde{w}^*) > \phi_1(\hat{w}) = e^{**}$.■

Proposition 2 demonstrates the Samaritan's dilemma outcome in our principal-agent context – the presence of the altruist affects the optimal wage contract and distorts the agent's incentives which leads to a reduction in effort and expected output. Intuitively, the presence of the altruist and the "safety net" he provides reduces the incentives for the agent to supply effort given that his punishment upon failure is effectively weakened.

3.2 A Non-Discriminating Altruist

Suppose now that the altruist cannot observe or verify whether the agent has failed or succeeded. Thus, a successful agent still has the possibility of consuming X (which determined his reservation utility before) but he can also claim failure and get the transfer $Y > 0$ from the altruist. Hence, the principal can obtain the output only if he can compensate the agent sufficiently.[6] Assuming that the principal is perfectly aware of this possibility, the participation constraints of the agent now become:

$$\begin{aligned} u(w_H) &\geq u(X+Y) \\ u(w_L) &\geq u(Y) \end{aligned}$$

As before, the low-output participation constraint always binds, i.e. $w_L = Y$, the ICC binds, but the high-output participation constraint may not. The expression determining whether the latter would occur is now

$$p'(\phi_2(X+Y))\phi_2'(X+Y)(V-X) - p(\phi_2(X+Y)) \tag{10}$$

[6] The agent cannot possibly continue with his contractual obligations with the principal and then collect the Y from the altruist, because the altruist now knows for sure that there was success, since otherwise the principal would not have catered to the agent.

where ϕ_2 is defined as before from (8). Let the optimal effort in this case be $e^{***}(w_H)$.

Proposition 3

(a) The agent's effort level with a non-discriminating altruist, e^{***} is higher than the effort with a "safety net" altruist but lower than in the case without an altruist, $e^{**} < e^{***} \leq e^{*}$.

(b) If the agent's participation constraint is binding and he is risk neutral, the presence of the non-discriminating altruist does not distort agent's effort i.e. $e^{***} = e^{*}$ – the Samaritan's Dilemma problem is avoided.

Proof:

(a) Proceed the same way as in Proposition 2. Given that ϕ_2 is increasing, note first that (10) being negative is implied by (9) being negative. Thus, the following three cases are possible:

(1) The PC for the safety net case binds, i.e. $w_H^{**} = X$ which implies $w_H^{*} = X$ and $w_H^{***} = X + Y$. Then we have that e^{***} is lower than e^{*} because $u(X+Y)-u(Y) \leq u(X) = u(X)-u(0)$ by the concavity of u (remember $w_L^{*} = 0$). Similarly, we have $e^{***} > e^{**}$ because, given $w_L^{**} = Y$, we have $u(X+Y)-u(Y) > u(X) - u(Y)$.

(2) the PC for the safety net altruist case does not bind but that for the non-discriminating altruist does, i.e. $X + Y > \tilde{w}^{**} > X$. Then it is clear from (8) that $e^{***} > e^{**}$. On the other hand, $e^{***} \leq e^{*}$ because $u(X + Y) - u(Y) \leq u(X) \leq u(\tilde{w}^{*})$ by the fact that u is increasing and concave.

(3) the PCs for both the safety net and the non-discriminating cases do not bind. Then $e^{**} = e^{***} = \phi_2(\tilde{w}^{**}) \leq e^{*}$ as shown in Proposition 2.

(b) In the case of a risk-neutral agent we have $u(X+Y) - u(Y) = u(X)$ and so, if the PC for the non-discriminating case binds, we have $e^{***} = e^{*}$.■

Interestingly, the fact that the altruist is unable to condition his aid upon output failure changes the optimal contract in a way that raises the effort level supplied by the agent relative to the situation where he could choose to provide the transfer upon failure only. Intuitively, this happens because the agent's incentives are now better aligned and the utility differential between success and failure is larger relative to the safety net altruist setting.

An important implication of this result is that the altruist – even if she can observe project outcome – might knowingly permit[7] such seemingly opportunistic behavior by the agent, since this makes the principal change the optimal wage contract accordingly in what turns out to be an efficiency enhancing way. Still, note that compared to the case without altruist, the agent generally supplies less effort (although clearly obtains higher utility). In the case of a risk neutral agent and binding participation constraints, however, the presence of an undiscriminating altruist does not distort incentives and efficiency at all and $e^{***} = e^{*}$ – the Samaritan's dilemma problem is completely avoided.

4 Endogenous Transfer from the Altruist

Until now we have treated the transfer, Y that the altruist provides to the agent as exogenous. In this section we endogenize the behavior of the altruist and derive the optimal transfer. Denote the endowment of the altruist by ω. After the project outcome is realized, the altruist decides how much of his endowment to transfer to the agent[8], $Y \in [0, \omega]$. The altruist consumes the residual amount, $\omega - Y$. Assume that the altruist's utility, U_{Alt} is

$$U_{Alt} = W(\omega - Y) + \alpha U_A$$

[7] Generally, that would depend on the cost of transfers and the altruist's objective function. See the next section for an extension to the model that endogenizes the altruist's transfer decision.

[8] As standard in the literature, we impose a non-negativity constraint on the transfer.

where the parameter $\alpha \geq 0$ represents the degree of altruism, the utility of the agent in the event of a failure[9] is U_A, and the function W is increasing and concave.

4.1 Safety Net Altruist

In this case, the altruist's maximization problem is:

$$\max_{Y} \; W(\omega - Y) + \alpha U_A$$

The altruist can perfectly observe the project outcome and hence $U_A = u(Y)$. Denote the solution to the above problem by $\bar{Y}$.

4.2 Non-Discriminating Altruist

In this case the altruist cannot observe the project outcome. Thus, if an agent comes to him, the altruist cannot tell for sure whether there was success or failure. The altruist takes this factor into consideration in his maximization problem which becomes:

$$\max_{Y} \; W(\omega - Y) + \alpha\{p(e)u(X + Y) + (1 - p(e))u(Y)\}$$

Denote the solution by Y^*. The optimal transfer Y^* that the altruist gives the agent is now a function of the agent's effort level. This is in contrast to the safety net altruist case in which the transfer $\bar{Y}$ is not contingent on the agent's effort level. Therefore, to compare the equilibrium effort levels, we must characterize how the optimal transfers compare in the two cases.

Proposition 4:

(a) For a risk-averse agent the optimal transfer that the non-discriminating altruist gives is strictly smaller than that of the safety net altruist, i.e. $Y^* < \bar{Y}$*.*

[9] We assume that the altruist cares about the agent only in case of a failure and not otherwise. Thus he provides transfers only to agents claiming adverse circumstances. This assumption is to hone in on the main purpose of the Samaritan's mission.

The agent's effort is larger in the non-discriminating altruist case compared to the safety net case.

(b) For a risk-neutral agent, we have $\bar{Y} = Y^*$ and the agent's effort in the non-discriminating altruist case equals the level when there is no altruist.

Proof: (a) Suppose not, i.e. $Y^* \geq \bar{Y}$. From the FOC of the altruist's problem we know that Y^* solves the following equation:

$$W'(\omega - Y^*) = \alpha[p(e)u'(X + Y^*) + (1 - p(e))u'(Y^*)]$$

Similarly, $\bar{Y}$ satisfies the equation

$$W'(\omega - \bar{Y}) = \alpha u'(\bar{Y})$$

If $Y^* \geq \overline{Y}$, then by the properties of W, we have:

$$\alpha[p(e)u'(X + Y^*) + (1 - p(e))u'(Y^*)] = W'(\omega - Y^*) \geq W'(\omega - \bar{Y}) = \alpha u'(\bar{Y})$$

or, simplifying:

$$p(e)[u'(X + Y^*) - u'(Y^*)] + u'(Y^*) \geq u'(\bar{Y})$$

By the concavity of u, given $X > 0$, the first term is negative. Also, if $Y^* \geq \bar{Y}$, we have $u'(Y^*) < u'(\bar{Y})$, so the right hand side of the above inequality must be larger than its left hand side – a contradiction. Thus $Y^* < \bar{Y}$.

For the second part of the proposition statement, note that from the agent's FOC we have: $u(w_H + Y^*) - u(Y^*) = \frac{g'(e)}{p'(e)}$. The result about effort levels then follows similarly to that in Proposition 3 since $Y^* < \bar{Y}$ and u is increasing and concave. Hence $e^{**} > e^{***}$ still holds as before.

(b) If the agent is risk-neutral, then u' is a constant and so the altruist's FOCs are the same in both the safety net and non-discriminating cases. Thus, $\bar{Y} = Y^*$. The result about agent's effort then follows exactly as in Proposition 3.■

The intuition for the result in Proposition 4 is that, when the altruist knows that the agent might come to him with positive probability even if he has been successful, he pays less to the agent due to the incomplete information problem and as a result the agent's effort level is higher. In addition, part (b) shows that the result we obtained in Proposition 3 whereby for a risk-neutral agent the Samaritan's Dilemma inefficiency result can be avoided remains robust when the altruist's transfer is endogenous.

5 Conclusion

The existing literature on the Samaritan's Dilemma problem has argued about the negative consequences on agents' effort or saving's behavior in situations were an altruist intends to help out the agent in bad states of the world. This result has important policy implications and consequently researchers have been looking for means to tackle the inefficiency problem.

In this paper we embed the Samaritan's Dilemma problem in a principal-agent framework. We find that the introduction of an altruist into a principal-agent setting might not necessarily have a negative effect on agent's effort as previously claimed. In fact, we demonstrate that if the agent is risk-neutral, the incentive problem is resolved entirely. This result generalizes to the case when the transfer from the altruist is endogenously determined. Our model allows for the possibility that the agent might consume the output and then take the handout from the altruist. An important point we make is that the altruist might actually support such seemingly opportunistic behavior since this helps attaining a more efficient outcome. The potential empirical and policy implications of our theoretical findings remain topics for future research.

References

[1] Buchanan, J. (1975), "The Samaritan's Dilemma", in *Altruism, Morality and Economic Theory*, edited by E. Phelps, Russel Sage, NY.

[2] Bruce, N and M. Waldman (1990), "The Rotten Kid Theorem Meets the Samaritan's Dilemma", *Quarterly Journal of Economics*, 105: 155-165.

[3] Bruce, N and M. Waldman (1991), "Transfers in Kind: Why They Can Be Efficient and Non-Paternalistic", *American Economic Review*, 81: 1345-1351.

[4] Coate, S. (1995), "Altruism, the Samaritan's Dilemma and Government Transfer Policy", *American Economic Review*, 85: 46-57.

[5] Ghatak, M., M. Morelli and T. Sjostrom (2001), "Occupational Choice and Dynamic Incentives", *Review of Economic Studies,* 68: 781-810.

[6] Ghosh, P., D. Mookherjee and D. Ray (2001), "Credit Rationing in Developing Countries: An Overview of the Theory" in *Readings in the Theory of Economic Development*, edited by D. Mookherjee and D. Ray, Blackwell.

[7] Hansson, I and S. Charles. (1991),"Social Security as Trade Among Living Generations", *American Economic Review*, 79: 1182-1195.

[8] Haveman, R., and B. Wolfe (1995), "The Determinants of Children's Attainments: A Review of Methods and Findings", *Journal of Economic Literature,* 33, 1829-1878.

[9] Karaivanov, A. (2007), "Financial Constraints and Occupational Choice in Thai Villages", working paper, Simon Fraser University.

[10] Kotlikoff, L. (1987), "Justifying Public Provision of Social Security", *Journal of Policy Analysis and Management*, 6:674-689.

[11] Lindbeck, A and J. Weibull (1988), "Altruism and Time Inconsistency: The Economics of Fait Accompli", *Journal of Political Economy*, 96: 1165-1182.

[12] Lagerlof, J., (2004), "Efficiency Enhancing Signalling in the Samaritan's Dilemma", *Economic Journal*, 114: 55-68.

[13] O'Connell, S. and S. Zeldes, (1993), "Dynamic Efficiency in the Gifts Economy", *Journal of Monetary Economics*, 31: 363-379.

[14] Rogerson, R., (1985), "The First-Order Approach to Principal-Agent Problems", *Econometrica,* 53(6): 1357-1367.

6 Appendix

Proof of Proposition 1:

We will show that the sign of the expression in (5) determines whether the high-output participation constraint, (3) binds. The principal's problem (given $w_L = 0$) is:

$$\max_{w_H \geq X,\, e} \; p(e)(V - w_H)$$
$$s.t.\; p'(e)u(w_H) = g'(e)$$

From the constraint, express $e = \phi_1(w_H)$ and plug into the objective function to obtain:

$$\max_{w_H \geq X} \; p(\phi_1(w_H))(V - w_H)$$

with a FOC

$$p'(\phi_1(w_H))\phi_1'(w_H)(V - w_H) - p(\phi_1(w_H)) \leq 0$$

with equality at an interior solution, $w_H > X$.

Clearly, the relevant range for w_H is $[0, V]$. Evaluated at $w_H = 0$, the above derivative is strictly positive, i.e., by continuity, profits are increasing (from zero) in w_H for low wage values. On the other hand, evaluated at $w_H = V$, the derivative equals $-p(\phi_1(V)) < 0$ i.e., by continuity, profits are decreasing in w_H for high values of the wage. Therefore, if profits are maximized at some interior value of w_H, it satisfies:

$$p'(\phi_1(w_H))\phi_1'(w_H) = \frac{p(\phi_1(w_H))}{V - w_H} \tag{11}$$

Lemma 1 *Given our assumptions on the functions $u(.)$, $p(.)$ and $g(.)$, the function $\phi_1(w)$ is strictly increasing in w.*

Proof of Lemma 1: Note first that $F(e) \equiv \frac{g'(e)}{p'(e)}$ is strictly increasing in e, since $F'(e) = \frac{g''p' - p''g'}{(p')^2} > 0$ because g is convex and p is concave. Thus, F is invertible and we can write $\phi_1(w) = F^{-1}(u(w))$. We then have $\phi_1'(w) = \frac{u'(w)}{F'(.)} > 0$ so ϕ_1 is strictly increasing.□

Lemma 1 implies that the RHS of (11) is an increasing function of w_H, starting from 0 at $w_H = 0$ and increasing towards $+\infty$ as $w_H \to V$. The LHS is positive over the interval $[0, V]$ and, given Assumption A1, it is monotonically decreasing by the concavity of $p(\phi_1(w_H))$[10] over this interval. Thus, the properties of the functions in the RHS and LHS imply that (11) has a unique interior solution, $\tilde{w}^* \in (0, V)$. Consequently, the profits for the principal are increasing for $w_H \in [0, \tilde{w}^*)$ and decreasing for $w_H \in (\tilde{w}^*, V]$.The latter directly implies that if (5) holds[11], then $\tilde{w}^* \leq X$ and hence $w_H^* = X$, while, if the opposite inequality is true, then $w_H = \tilde{w}^* > X$. Q.E.D.■

[10] A sufficient (but not necessary) condition for this is $F'' > 0$.
[11] Note that (5) is simply (11) evaluated at X.