Australian School of Business Research Paper No. 2013 ECON 21

Understanding social impact bonds and their alternatives: An experimental investigation

Jade Wong
Andreas Ortmann
Alberto Motta
Le Zhang

UNDERSTANDING SOCIAL IMPACT BONDS AND THEIR ALTERNATIVES: AN EXPERIMENTAL INVESTIGATION *

Jade Wong, Andreas Ortmann**, Alberto Motta, Le Zhang
Australian School of Business, UNSW, Sydney

September 9, 2013

*We thank seminar audiences at Monash University, UNSW, CERGE-EI, University of Heidelberg, University of Amsterdam, and WissenschaftsZentrum Berlin for excellent questions and suggestions. We are particularly grateful to Chris Bidner for an excellent discussion of an earlier version of the present manuscript.
**Corresponding author:
Via email: a.ortmann@unsw.edu.au, aortmann@gmail.com
Via mail: Room 460, Australian School of Economics, University of New South Wales Sydney, NSW 2052 (Kensington), Australia

ABSTRACT

Policy-makers world-wide have proposed a new contract – the "social impact bond" (SIB) – which they claim can allay the underperformance and underfunding afflicting not-for-profit sectors, by tying the private returns of (social) investors to the success of social programs (Bolton 2010; Bolton & Savell 2010; Mulgan et al. 2010a,b; Liebman 2011; Tierney & Fleishman 2011; Von Glahn & Whistler 2011). Given the high hopes governments on various levels in England, Australia, and New York have pinned on this contract format, the considerable amount of money that has recently been poured into this emerging market (e.g., http://www.bigsocietycapital.com/), and the fact that serious are program evaluations cannot be expected any time soon (Disley et al. 2011; see also McKay 2013 and Pratt 2013), we test this new contract by way of experimental methods. We report an investigation of how SIBs perform in a first-best world, where investors are rational and able to obtain hard information about not-for-profits' performance. To this end, we use a principal-agent multi-tasking framework to compare SIBs to inputs-based (IBs) and performance-based (PBs) contracts, which represent the most commonly used contracts governments and not-for-profits write. IBs contain a piece-rate mechanism, PBs contain a non-binding bonus mechanism, and SIBs contain a mechanism that, due to the presence of an investor, offers full enforceability. Although SIBs can perfectly enforce good behavior, they also require the principal (i.e. government) to relinquish control over the agent's (i.e. not-for-profit's) payoff to a self-regarding investor, which prevents the principal and agent from being reciprocal. In spite of these drawbacks, in our experiment SIBs outperformed IBs and PBs. We therefore conclude that, at least in our laboratory test-bed, SIBs can indeed allay underperformance and therefore possibly underfunding of not-for-profits.

INTRODUCTION

Not-for-profit sectors world-wide account for 5 – 10 percent of GDP; providing a variety of quality-adjustable goods and services that often contribute to civil society (Powell & Steinberg 2006; Anheier & Toepler 2010). Not-for-profits are also, to various degrees, plagued with problems of underperformance, such as overspending on administration and fundraising, mismanaging funds, and engaging in outright fraud (e.g., Ortmann & Schlesinger 2003; Kristof 2011; Rusche & Burke 2011; BDO 2012; NVCO 2011).

To the extent that not-for-profits are funded significantly by governments, they have a special fiduciary duty to society to deliver their services efficiently. In Australia, for example, half of not-for-profits' revenues are obtained through various forms of funding from federal and state governments (Productivity Commission 2010; henceforth PC, 2010). Unfortunately the current modes of contracting, relying overwhelmingly on what we call inputs-based and performance-based contracts, and social impact measurement are far from perfect (PC 2010). The problem seems to be two-fold: on the one hand, not-for-profits are rarely as transparent and accountable as society has a right to expect; on the other hand, governments are both bad at monitoring because they might not have the technology or the funds to spend on monitoring, and unreliable in their long-term provision of funds (e.g., PC 2010; see also Boris et al. 2010). These factors lead to underfunding. This underfunding problem, in addition to the problems of tight public budgets and bureaucrats' hesitation to undertake risky projects, has provided the ground for the enthusiastic reception of social impact bonds (e.g., Bolton & Savell 2010; Liebman 2011; Mulgan et al. 2010; Roth 2011; Von Glahn & Whistler 2011; Tierney & Fleishman 2011; or www.socialfinance.org.uk, www.social-finance.com.au).

Social impact bonds introduce a new player to the typical government-not-for-profit provider contractual relationship - the social investor. This investor has social preferences and, under other circumstances, might have donated to some worthy cause but has decided that her or his dollar is better spent on giving funds to worthy – albeit risky – social projects, preferably through suitably diversified social innovation funds. The details of the funding do not matter for the present purpose, but we do assume that these investors have greater incentives and are indeed better equipped than government bureaucrats to monitor the not-for-profit provider, since their financial return depends on social return. We also assume that investors are willing to take on risks that bureaucrats might shy away from, because they can suitably diversify away idiosyncratic risks.

To provide context, consider the Social Impact Bond (SIB) that was piloted first, and is currently underway at the UK's Peterborough Prison. In 2010, the UK Justice Ministry contracted the social investment bank, Social Finance, to raise capital from outside investors to fully fund the cost of not-for-profits' in-prison and community intervention programs (for details, see Social Finance 2011 and Disley et al. 2011). The agreed aim was to reduce the re-offending rate of 3000 short-sentence male prisoners by 7.5% or more over 6 years (Bolton 2010). If re-offending decreases by at least 7.5%, the government repays Social Finance and thus the investors the cost of funding, in addition to interest. [1] If the 7.5% threshold is not reached, Social Finance and its investors lose their entire investment, which includes the principal and return (Liebman 2011).[2]

The basic timing of the transactions involved in SIBs is thus as follows: First, (social) investors enter a contractual relationship with the government, where they agree to fully fund a not-for-profit to deliver a program that aims to improve the prospects of a target group. Second, the not-for-profit uses investors' funds to implement the program, which is eventually evaluated. Lastly, the government repays investors the cost of funding the program plus interest, where repayments are tied to levels of improvement in the target group.

In a first-best world, investors are rational and able to obtain "hard" monitoring information about the not-for-profit's performance. Hard monitoring renders the information credible in a court of law, and so truthfully reveals the not-for-profit's performance (Tirole 1986). If these assumptions hold, SIBs offer "perfect enforceability" – it allows investors to perfectly write a contract on the not-for-profit's performance. The advantage of perfect enforceability is countered by the government's loss of control. Specifically, by delegating the contracting task to the investor, the government is no longer able to influence the not-for-profit's behaviour or payoff; for example, the government might not be able to draw on reciprocal behaviour that might exist based on conceptualizations of not-for-profits as organizational entities *sui generis* (see, for a discussion and references, Ortmann 1996).

[1] Preliminary figures for this project, although not conclusive, suggest that it might work: The Ministry of Justice has reported a decline of six percentage points in the frequency of reconvictions, "from 87 in the period 2008 to 2010 to 81 in the period 2010 to 2012. The compares with a 10 percentage points nationally over the same periods, from 69 in 2008 to 2010 to 79 in 2010 to 2012." (Pudelek 2013)

[2] The way this particular bond is structured is obviously problematic. While it completely shifts risk away from the government, it imposes potentially a very high penalty on investors for not reaching the target. This, unfortunately, is likely to mean that investors will only agree to "safe" targets, something that seems not desirable from a societal welfare point of view. We do not consider these important design issues in the present manuscript.

With this trade-off in mind, we test the efficacy of SIBs in a laboratory setting. Our experiment reflects the multi-task principal-agent nature of the interactions of governments and not-for-profits, as it reflects the type of goods (i.e. "experience goods") not-for-profits typically provide (Holmstrom & Milgrom, 1991; Ortmann 1996; Ortmann & Schlesinger 2003). Specifically, the government (neutral) offers the not-for-profit (masculine) a contract, which he accepts or rejects. If he accepts the contract, he engages in a chosen-effort task. He chooses effort on task 1 (a quantitative task) and task 2 (a qualitative task), which is costly for the not-for-profit but increases the government's payoff. We assume the government can perfectly observe the not-for-profit's effort on both tasks, but can only verify in a court of law the effort on task 1. This set-up thus reflects governments' inability to properly monitor in real life (e.g. see PC 2010).

We begin our exploration by investigating the pitfalls of two widely used contracts government currently offer not-for-profits - inputs-based (henceforth, IBs) and performance-based contracts (henceforth, PBs). IBs contain a wage and a piece-rate, which is paid for every unit of effort the not-for-profit exerts on task 1. PBs contain a wage and the promise of a bonus. The bonus might be paid once the government observes the not-for-profit's chosen effort levels, but the payment is not enforceable due to the unverifiable nature of task 2. If governments and not-for-profits are payoff maximisers and interactions are one-shot, PBs are theoretically predicted to fail. Governments are theoretically predicted to never pay the promised bonus, and so not-for-profits, backward inducting, will exert minimum effort on tasks 1 and 2. In contrast, governments can induce not-for-profits to exert effort on task 1 through the piece-rate component in IB.

From a societal (and government's) viewpoint, the theoretical predictions are that IBs are preferred to PBs but are dominated by SIBs. In our experiment, SIBs outperformed IBs and PBs. Both IBs and PBs performed poorly, but PBs outperformed IBs, as – driven by reciprocity -- they elicited higher and more efficient effort levels. Notwithstanding widespread government cheating in PBs, government payoffs and social surplus were also higher in PBs.

The rest of this paper is as follows: Section 2 expands on the motivation and the literature. In Section 3, we formulate our bare-bones multi-effort principal-agent model, which guides our experimental exploration of the three contract formats. Section 4 details the experimental design and implementation, and Section 5 presents our results. We discuss the main results in Section 6, and provide concluding remarks and future research avenues in Section 7. An

appendix provides information about a control treatment (Appendix 1), instruction materials (Appendix 2), our calibration exercise (Appendix 3), and further demographic information (Appendix 4).

2. A REVIEW OF THE LITERATURE

We review the literature in three subsections to motivate our study: The nature of not-for-profits (i.e., the problem of underperformance); the institutional setting in which they operate (i.e., the credibility problem resulting from underperformance, and the problem of underfunding that underperformance triggers), and the mechanics of social impact bonds, which many believe are able to address these problems (e.g., Disley et al. 2011).

2.1 NOT-FOR-PROFITS

Contrary to widespread opinion, not-for-profits are allowed to make profits. They are, however, not allowed to distribute profits to their "owners" or controlling parties.[3] Hansmann (1980, 1987) has argued that this "non-distribution constraint" is an effective means of preventing opportunism. Ortmann (1996), Ortmann & Schlesinger (2003), and others have pointed out that the "non-distribution constraint", even in conjunction with the "reasonable compensation constraint" (the constraint that salaries have to be in line with comparables), leaves considerable leeway for opportunistic behaviour due to the specific nature of goods and services many not-for-profits provide.

These goods and services – often social services such as day-care, old-age care, health, and education services – are known as "experience goods", because their quality cannot be evaluated by customers at the time of purchase, or even after experience (in which case they are called "credence goods"). When customers cannot assess the quality of the good delivered, opportunistic not-for-profits may exploit their information advantage by delivering a lower quality than promised. By virtue of the non-distribution constraint and the reasonable compensation constraint, incentives to behave opportunistically are allegedly blunted because revenue that accrues from short-changing customers cannot and do not benefit the owners of not-for-profits.

Ortmann (1996), drawing on Klein & Leffler (1981) and Kreps (1990), shows that the inner dynamics of not-for-profits and their interaction with customers can be conceptualized by

[3] They can, however, distribute them to a subset of their "stake-holders", namely those they serve.

simple reputation or principal-agent games. These games theoretically lead to socially inefficient equilibria in one-shot and finitely repeated games, which in our case is when not-for-profits provide low quality goods and governments do not trust them to do so.[4] Yet the folk theorem posits that if players are sufficiently patient, the shadow of the future is sufficiently long, and perfect information exists, providing high quality goods and trusting not-for-profits is a sustainable equilibrium, as reputation enforcement – where not-for-profits with good track records attract more customers, or government support – will induce not-for-profits to not cheat. Ortmann & Schlesinger (2003) discuss to what extent these conditions apply to various kinds of not-for-profits. They argue that the dearth of transparency in the not-for-profit sector and the difficulties associated with verifying the quality of experience goods makes reputational enforcement difficult to achieve because the condition of perfect, or at least adequate, information is rarely satisfied. Hence, the dismal outcome where not-for-profits opportunistically underperform by providing poor quality goods and services remains a real problem. It is this potential for underperformance that causes the credibility problem faced by not-for-profits. Donors, for example, will only give to not-for-profits if they can reasonably trust that their donations will go to cause (Wong & Ortmann, 2012a).

2.2 INSTITUTIONAL SETTINGS

Governments are increasingly reliant on not-for-profits to deliver experience goods such as housing, intervention programs and employment training (Boris et al. 2010; Erwin 2011; PC 2010). As a consequence, many not-for-profits are heavily dependent on government funding. For instance, around 20,000 Australian not-for-profits relied heavily on government funding during the 2006-07 period (The Treasury 2011), and almost 33,000 American not-for-profits received government funding in 2009; it was the largest source of revenue for 60% of them (Erwin 2011). This reliance on public funds imposes a special fiduciary duty to society for not-for-profits to deliver services efficiently. However the ways in which not-for-profits actually use public funds to create value is poorly understood.

It seems reasonable to say that the current state of monitoring and evaluation is underwhelming. PC (2010) surveyed 43 Australian not-for-profits about 109 of their programs. While 97% of programs employed performance reporting, the majority was conducted in-house rather than independently (refer to D.13 of PC 2010). Similarly, American not-for-profits that earn revenue over US$25,000 are required to complete an IRS

[4] The behavioural evidence in favour of that proposition, however, is to some extent contested.

990 form detailing their financial performance, but the information is self-reported and does not require external appraisal. Self-assessment is problematic on many levels. First, many not-for-profits do not possess the correct monitoring infrastructure to collect and evaluate performance data, which complicates monitoring and reduces the reliability of the information (PC 2010). Second, even when proper monitoring infrastructure exists, not-for-profits can fabricate self-reported performance information (e.g. see Krishman et al. 2006; Yetman & Yetman 2011b; Friesen & Gangadharan 2012a, 2012b). Lastly, government departments may not be able to evaluate not-for-profits' performance information adequately, since they typically have low-powered incentives and questionable performance measures (PC 2010; Liebman 2011).

In response to the problems associated with poor monitoring, government contracts are overly prescriptive, specifying how funds should be used and what services should be provided (PC 2010). An example is the "purchase of services" contract, which is where not-for-profits deliver governments' desired inputs and/or outputs in exchange for money. While the performance-based aspect of the contract is theoretically appealing, its effectiveness hinges on two conditions. First, governments must be able to measure the contracted inputs and/or outputs accurately and objectively. However this is difficult in light of the underwhelming state of monitoring already alluded to, and the experience nature of the goods not-for-profits typically provide. Second, the contracted inputs and/or outputs must correlate to social value. If it does not, not-for-profits might place most of their effort on increasing the contracted inputs/outputs rather than fulfilling their mission (Holmstrom & Milgrom 1991).

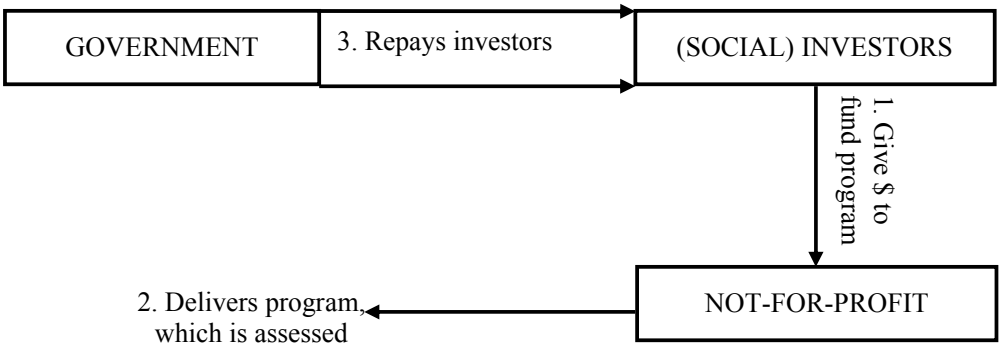
2.3 SOCIAL IMPACT BONDS

In light of the current not-for-profit landscape, policy-makers designed a new contract form, "Social Impact Bonds" (SIBs). Mulgan et al (2010a) generalize SIBs' basic properties in three steps (refer to Figure 2.3.1, next page). First, (social) investors enter a contract with the government, where they agree to fully fund a not-for-profit to deliver a program which aims to improve the livelihoods of a target group. Second, the not-for-profit uses investors' funds to deliver the program, which is then assessed. Third, the government repays the investors the cost of funding the program with interest, and concurrent with the level of improvement in the target group.

The Petersborough Prison funding scheme discussed in the introduction illustrates this scheme. SIBs shift the risk for such interventions away from bureaucrats to investors, and

give investors more incentive to monitor and evaluate the performance of the service provider. The "pay-for-success" feature of social impact bonds gives them considerable equity characteristics, even though they are called bonds.

FIGURE 2.3.1: SOCIAL IMPACT BONDS



Advocates describe social impact bonds as a win for all sectors involved. The public sector benefits because the risk of failure is borne, depending on the design of the SIB, fully or partially by investors. The private sector benefits because social impact bonds give investors an opportunity to invest in both social and financial returns. The not-for-profit sector benefits because they receive full-cost-of-funding upfront, and are not subject to prescriptive government contracts.

3. THEORETICAL FRAMEWORK AND PREDICTIONS

We construct three bare-bones models that capture the three contract formats governments use to fund not-for-profits: inputs-based (IB), performance-based (PB), and social impact bond (SIB) contracts. We also construct a first-best contract for comparison purposes (see Table 3.0.1).

TABLE 3.0.1: GOVERNMENT CONTRACTS

First-Best Contract	**Inputs-Based Contract (IB)**
The government writes a contract on outcomes; the contract is legally enforceable	The government pays the not-for-profit for every verifiable input placed in the program.
Performance-Based Contract (PB)	**Social Impact Bond Contract (SIB)**
The government promises to pay the not-for-profit once it achieves a pre-specified performance target. Payment is not legally	An investor funds the not-for-profit. The government pays the investor if the not-for-profit achieves a pre-specified performance

enforceable.	target. Payment is legally enforceable.

To motivate the models, consider the following story inspired by the Peterborough pilot. The government is concerned about the high level of prisoner recidivism, as it entails high societal costs. The government thus seeks ways to rehabilitate prisoners. It contracts a not-for-profit to deliver rehabilitation services (henceforth "program"). Government cost savings depend on the number of prisoners that participate in the program and the quality of the program delivered. Though the government can verify the number of prisoners that participate in the program, it finds it more difficult to verify the quality of the program, as it lacks the proper monitoring technology and incentives to do so. Government bureaucrats might also shy away from the types of social innovation that can go awry and then impose significant (reputational) costs on them.[5] In light of the government's limitations, it asks: what contract should it implement?

3.1 BASIC SET-UP

Assume initially two players: the government (neutral) and the not-for-profit (masculine). Government cost savings, Se_1e_2, depend on the level of effort the not-for-profit exerts on tasks 1 and 2, e_1 and e_2. Task 1 entails quantitative elements, such as enrolment in classes. Task 2 entails qualitative elements, such as various interventions specifically tailored towards each prisoner's needs. The functional form of cost savings reflects the assumption that for any given effort level, government cost saving is maximised when effort is equally exerted on task 1 and 2. [6] The government's budget is its cost savings minus the remuneration ("wage", for short) it pays the not-for-profit.[7]

$$B_{government} = Se_1e_2 - w$$

The not-for-profit's cost of delivering the service is $c(e_1,e_2)$ and depends on the total amount of effort he exerts on tasks 1 and 2. Cost of effort is increasing and convex ($c_{e1} > 0$, c_{e1e1} and

[5] Relatedly, governments have various options of how to finance educational offerings: They can make financial support (grants, scholarships, direct funding to educational institutions) dependent on easily quantifiable key performance parameters such as enrollment in classes, or qualitative components such as individual consultations that are much more difficult to evaluate, and possibly only with large delays.

[6] This strong complementarity seems reasonable for the kind of contexts that we consider. Obviously, it might be desirable to relax this assumption to some extent in other contexts.

[7] The government's objective is to maximise cost savings. This makes sense because they allegedly seek the "best value for money" when they offer contracts to not-for-profits (PC 2010:297), and because SIBs were conceived to generate "greater taxpayer's value for money", which implies governments are concerned about efficiency.

c_{e2}, $c_{e2e2} > 0$), and the lowest effort level he can exert on tasks 1 and 2 is 1, where $c(1,1) \neq 0$. The not-for-profit's "utility" equals his wage minus the cost of effort.

$$U_{\text{not-for-profit}} = w - c(e_1,e_2)$$

Crucially, we assume the government is unable to monitor the not-for-profit's performance in a satisfactory manner, as it lacks the proper incentives and technology to do so. Hence, although the government can observe the level of effort the not-for-profit exerts on tasks 1, it can only verify effort on quantifiable task 1 in a court of law.[8] It cannot verify effort on qualitative task 2.

The timing of events is as follows: At date 0, the government makes a take-it-or-leave-it offer to the not-for-profit. If the contract is accepted, at date 1 the not-for-profit exerts effort on tasks 1 and 2. If the contract is rejected, both players receive their reservation utilities of zero. Lastly, payoffs are realised.

3.2 BENCHMARK CONTRACT (FIRST-BEST)

In the benchmark contract, the government is able to costlessly verify effort on tasks 1 and 2 in a court of law. It offers the not-for-profit a contract $\{e_1^*, e_2^*, w\}$, which specifies its desired effort levels on tasks 1 and 2, e_1^* and e_2^* and wage, w. The not-for-profit receives $w > 0$ if $\{e_1, e_2\} = \{e_1^*, e_2^*\}$ and $w = 0$ if otherwise. The government thus maximises its budget subject to the not-for-profit's participation constraint,[9] which yields first-best effort levels:

$$\{e_1^*, e_2^*\} = \text{argmax}_{e1,e2}\, Se_1e_2 - c(e_1, e_2)$$

First-best effort levels are achieved when the government's marginal cost savings equal the not-for-profit's marginal disutility of effort. The government offers the not-for-profit a wage $w^* = c(e_1^*,e_2^*)$ such that his participation constraint binds.

[8] To show the validity of the assumption that effort is observable but not verifiable, consider two examples: education and prisoner rehabilitation. A parent (principal) might be able to observe her or his child's attitude towards learning and learning outcomes but essentially the parent is unable to verify what happens in the classroom day-in day-out. Similarly, consultants are often paid to determine the quality of prisoner rehabilitation programs (e.g. see ARTD Consultants, 2007). While they can obtain information about "client satisfaction" through surveys, they often cannot, and do not, quantify the effect of the program on recidivism via rigorous methods, such as a randomised control experiment. Hence, while the government is in principle able to observe the amount of effort the not-for-profit service provider exerts in delivering the program, it cannot verify it.

[9] Formally, $\text{Max}_{e1,e2} Se_1e_2 - w$ subject to $w - c(e_1, e_2) \geq 0$

3.3 INPUTS-BASED CONTRACT (IB)

When the government cannot verify effort on task 2 in a court of law, it might offer an inputs-based contract {w, s}, where s is the level of compensation the not-for-profit receives for every unit of effort he exerts on task 1. To solve the contract, the government ensures the not-for-profit's participation constraint binds (1), and the contract is incentive compatible (2):

$$\text{Max}_{w,s}\ Se_1e_2 - w - se_1$$

Subject to

$$w + se_1 - c(e_1, e_2) \geq 0 \qquad (1)$$

$$\{e_1, e_2\} = \text{argmax}_{e1,e2}\, w + se_1 - c(e_1, e_2) \qquad (2)$$

The not-for-profit focuses his effort on quantitative elements (i.e. task 1) rather than on improving the quality of the program (i.e. task 2). He therefore exerts minimum effort on task 2, $e_2 = 1$, but increases effort on quantifiable e_1 until his marginal gain from receiving s equals his marginal disutility of effort $c_{e1}(e_1, 1)$. Given the not-for-profit exerts $e_2 = 1$ and e_1 such that $s = c_{e1}(e_1,1)$, the government implicitly sets optimal e_1^* such that it solves $S = c_{e1}(e_1^*,1) + c_{e1e1}(e_1^*,1)e_1^*$. Thus, optimal effort levels are achieved when the government's marginal gain from cost savings equals the marginal cost of funding the not-for-profit. Lastly, the government offers compensation $s^* = c_{e1}(e_1^*,1)$ and wage $w^* = c(e_1,1) - c_{e1}(e_1,1)e_1$ such that the not-for-profit receives his reservation utility.

3.4 PERFORMANCE-BASED CONTRACT (PB)

In performance-based contracts, the government offers contract {w, b*}, where b* is the promised bonus. The government also sets a performance target $\{e_1^*, e_2^*\}$, which specifies its desired effort on tasks 1 and 2. The government announces the not-for-profit will receive $b^* > 0$ if he satisfies the performance target $\{e_1^*, e_2^*\}$, and $b^* = 0$ if otherwise. However the non-verifiability of task 2 means the promise of a bonus b* is not legally enforceable in a court of law[10]. The payoffs for PBs are thus:

[10] Poor monitoring makes verifying the performance of not-for-profits difficult. In such cases, governments may unwittingly (or intentionally) renege on their promise to pay not-for-profits for their services. For instance, in 2009 58% of American human-services not-for-profits experienced late payments, payment cancellations or payment cuts to their contracts (Boris et al. 2010), which arguably are forms of governments "reneging" on promised payments. Similar data can be found in PC (2010, specifically D.13).

$$B_{government} = Se_1e_2 - w - b$$
$$U_{not\text{-}for\text{-}profit} = w + b - c(e_1,e_2)$$

The government never fulfills its promise to pay $b^* > 0$ when $\{e_1, e_2\} = \{e_1^*, e_2^*\}$, because it is budget maximising to set b equal to zero. When b = 0, the not-for-profit maximises his utility by exerting the lowest possible effort on tasks 1 and 2, $\{e_1, e_2\} = \{1, 1\}$. The government thus offers a wage $w^* = c(1,1)$ such that the not-for-profit's participation constraint binds.

3.5 SOCIAL IMPACT BOND (SIB)

Unlike government bureaucrats who face tight budgets (which can crowd out funding of monitoring activities), red tape, and want to avoid risk, investors have discretion over their funds; investors are also less constrained by administrative inefficiencies, and are better equipped to handle risk. An investor (feminine) in social impact bonds has better aligned incentives, resources and technology to obtain hard monitoring information about a not-for-profit's performance. Hard information prevents the investor from manipulating the monitoring report and the government from doubting it, as it credibly conveys the not-for-profit's performance[11]. The investor can thus verify and write a contract on the not-for-profit's effort on tasks 1 and 2. Monitoring costs M and does not require effort.

In SIBs, the government offers the investor $\{e_1^*, e_2^*, B, r\}$, which contains a performance target $\{e_1^*, e_2^*\}$, the value of the bond B, and the bond's interest rate r. If the performance target $\{e_1^*, e_2^*\}$ is achieved, the government pays the investor B at interest rate r, where $r \in [0, 1]$, and B = 0 if otherwise. The investor's reservation interest rate (r_o) is normalized to zero. We assume that the investor does not possess the skills to deliver the program but can outsource the tasks to a not-for-profit.

The investor offers the not-for-profit a contract $\{w, e_1^*, e_2^*\}$, where he receives wage $w > 0$ if he achieves the government's performance target $\{e_1^*, e_2^*\}$, and w = 0 otherwise.[12] By outsourcing the task, the investor's financial return depends on the not-for-profit's

[11] We assume away the moral-hazard problem that undoubtedly would exist if an investor herself would be allowed to do the monitoring. The question of who monitors the monitor is important but can not be dealt with here.

[12] This assumption reflects the fact that in SIBs, not-for-profits are paid in instalments. Hence, if not-for-profits underperform, investors may stop making financial payments, or withdraw their support altogether.

performance. Hence, if the performance target is achieved, the investor receives $B(1 + r)$ from the government. If not, the investor receives 0 and incurs a monitoring cost of M.

The government therefore designs a contract that maximises its budget and ensures the investor's and not-for-profit's participation constraints binds:

$$\text{Max}_{e1,\, e2}\, Se_1e_2 - B$$

Subject to:

$$w - c(e_1, e_2) \geq 0 \qquad (3)$$

$$(1 + r)B \geq (1 + r_o)B \qquad (4)$$

$$(1 + r)B \geq w + M \qquad (5)$$

Optimality requires that all participations constraints bind, hence $r^* = 0$, $w^* = c(e_1, e_2)$, and $B^* = c(e_1, e_2) + M$. The optimal efforts levels are thus:

$$\{e_1^*, e_2^*\} = \text{Max}\{\text{argmax}_{e1,e2}\, Se_1e_2 - c(e_1, e_2) - M, 0\}$$

Note that there is no distortion on optimal effort levels since M is a constant. Therefore if the government's participation constraint is satisfied (i.e. $Se_1e_2 - c(e_1, e_2) \geq M$) and the investor is able to obtain hard monitoring information about the not-for-profit's performance, SIBs induce first-best effort levels.

3.6 PREDICTIONS

The investor in SIBs is able to obtain hard monitoring information about the not-for-profit's performance. She can thus perfectly write a contract on the not-for-profit's effort on tasks 1 and 2, which means SIBs can induce first-best effort levels. As such, SIBs will outperform IBs and PBs. Likewise, IBs will outperform PBs, since the piece-rate component allows the government to write a contract on task 1. PBs will fare worst, since the bonus payment cannot enforce effort at all.

4. EXPERIMENT

The experiment is based on the models developed in section 3, and explores the interaction between governments and not-for-profits under various contractual alternatives.

4.1 EXPERIMENTAL DESIGN

The government (neutral) offers the not-for-profit (masculine) a contract, which he accepts or rejects.[13] If he accepts the contract, he chooses effort on tasks 1 and 2, which is costly for the not-for-profit but increases the government's payoff. As in the theoretical framework, the government can perfectly observe the not-for-profit's effort on both tasks, but can only verify effort in a court of law on task 1.

Government cost savings are $Se_1e_2 = 75e_1e_2$, where $e_1 \in [1, 6]$ and $e_2 \in [1, 6]$. The not-for-profit's cost of effort is shown in Table 4.1.1. The marginal cost of effort is 30 when $e_1+e_2 \in [2, 6]$, and increases to 75 when $e_1+e_2 \in [7, 12]$, to approximate in a simple fashion the convexity of the cost function.[14]

TABLE 4.1.1: COST OF EFFORT

$e = e_1+e_2$	2	3	4	5	6	7	8	9	10	11	12
$c(e_1, e_2)$	60	90	120	150	180	255	330	405	480	555	630

In SIBs, the investor was computerized and programmed to behave in a selfish and payoff-maximizing manner. This was publicly announced; hence, it can be assumed that it was common knowledge. To reflect these assumptions, we implemented the following:

1. The investor's monitoring cost is 50 points. In return, the investor can obtain hard monitoring information about the not-for-profit's performance and perfectly enforce effort on tasks 1 and 2. Thus, if the not-for-profit complies with the government's desired effort levels, the not-for-profit and investor receive their wage and bond payment respectively. If not, they both receive zero points.

[13] Governments and not-for-profits are not individual decision makers. Hence, we need to be aware that the participants – whether they act knowingly or unknowingly in their role of organizational entities – may bring to the experiment social preferences like altruism or fairness that organizational entities may not be afflicted with. There is a literature which addresses the question whether individual or team decision making leads to significant differences (e.g. see Kugler et al. 2007; Kocher & Sutter 2007; Luhan et al. 2009); our reading is that the answer to this question remains open and could in any case be tested if necessary.

[14] The experimental parameters, particularly the S parameter, were calibrated on rehabilitation data on juvenile delinquents in NSW (particularly Holmes 2011; Juvenile Justice NSW 2011; and New South Wales Criminal Court Statistics 2010), which is the target of the proposed SIB pilot in New South Wales (CSI 2011). Details on the calibration exercise are available upon request. Our motivation to calibrate is twofold. First, experimentalists aim to increase the external validity of their experiments (e.g. see List 2006). Calibrating our parameters brings our experiment one step closer to the real world. This is an important step towards "parallelism", or external validity, which is of importance in public policy situations (e.g., Plott 1987). Second, the way in which experiments are parameterized can affect behaviour in the laboratory (e.g. see Engelmann & Ortmann 2009; Ridgon 2002).

2. The investor extracts surplus from the not-for-profit. It pays the not-for-profit a wage $w = 200 + c(e_1^*, e_2^*)$ if he exerts the government's desired effort levels on tasks 1 and 2. Hence, the not-for-profit earns 200 points when he complies with the government's desired effort levels.
3. The investor always accepts the government's SIB offer if her reservation utility of zero is satisfied (i.e. if $B \geq 250 + c(e_1^*, e_2^*)$).[15]

Table 4.1.2, is based on the theoretical model developed in Section 3 and contains the exact steps and payoffs in each contract.

TABLE 4.1.2: STEPS AND PAYOFFS

	IB	**PB**	**SIB**	
0	G offers {w, s} and specifies $\{e_1^*, e_2^*\}$	G offers {w, b*} and specifies $\{e_1^*, e_2^*\}$	G offers $\{e_1^*, e_2^*, B\}$ to I	
1	N chooses e_1 and e_2	N chooses e_1 and e_2	I offers $\{e_1^*, e_2^*, w\}$ to N, where $w = c(e_1^*, e_2^*) + 200$	
2	-	G chooses b	N chooses e_1 and e_2, which I monitors for 50 points	
Payoffs	$E^G = 75e_1e_2 - w - se_1$ $E^N = w + se_1 - c(e_1,e_2)$	$E^G = 75e_1e_2 - w - b$ $E^N = w + b - c(e_1,e_2)$	If $\{e_1, e_2\} = \{e_1^*, e_2^*\}$ $E^G = 75e_1e_2 - B$ $E^I = B - 50 - w$ $E^N = 200$	If $\{e_1, e_2\} \neq \{e_1^*, e_2^*\}$ $E^G = 75e_1e_2$ $E^I = -50$ $E^N = -c(e_1,e_2)$

* E = Experimental Points; G = Government; N = Not-for-profit; I = Investor

Each session lasted at least 8 periods.[16] To mitigate reputation building, governments were randomly matched to not-for-profits. This was also publicly announced and is thus assumed to be common knowledge. Participants were matched anonymously, could only observe the outcomes of their matches, and could not observe the past actions of others. We used "loaded" language in the instructions, as we portrayed the relationship between government and not-for-profit as an employer-employee relationship.

[15] Since $r^* = 0$ in the theoretic framework, we remove it in the experiment.

[16] The length of the experiment depended on how long it took participants to make their decisions. The recruitment email said that the experiment might last up to two hours. Once the targeted time was about to lapse, we stopped a session even if the targeted ten rounds were not reached; since that happened in a couple of cases, and to simplify our exposition, we report throughout only the results for the first 8 periods. None of our results are affected by this truncation.

We conducted two treatments. In treatment **IP**, governments could offer either IBs or PBs to not-for-profits, to reflect the government formats that are currently used. In treatment **PS**, governments could offer either PBs or SIBs to not-for-profits.

4.2 PREDICTIONS

Table 4.2.1 contains the predictions for the experiment, using the standard assumption of payoff-maximisation. In IBs, the not-for-profit exerts the lowest possible effort on task 2, $e_2 = 1$, and so the government's marginal gain from increasing e_1 by one unit is 75. Since the marginal cost of effort is 75 when $e_1 + e_2 > 6$, the government sets $e_1^* = 5$, $e_2^* = 1$, $s^* = 30$ and $w^* = 30$ such that the not-for-profit's participation constraint binds. In PBs, pre-empting the government does not deliver its promised bonus (i.e. b = 0), the not-for-profit exerts the lowest possible effort on tasks 1 and 2, $e_1 = e_2 = 1$. The government therefore offers wage $w^* = 60$ such that the not-for-profit's participation constraint binds. In SIBs, the government specifies first-best effort levels on tasks 1 and 2, $e_1^* = e_2^* = 6$, to maximise its budget, and offers the investor a bond $B^* = 630 + 200 + 50 = 880$. The investor therefore offers the not-for-profit a wage $w^* = 200 + c(e_1^*, e_2^*) = 200 + 630 = 830$, which ensures the not-for-profit receives 200 points if he complies with the government's desired effort levels. Finally, we predict that SIBs will outperform IBs, and IBs will outperform PBs in terms of social surplus (i.e. the sum of all payoffs). Table 4.2.1. summarizes the predictions.

TABLE 4.2.1: SUMMARY OF PREDICTIONS

	IB	PB	SIB
(e_1, e_2)	(5,1)	(1,1)	(6, 6)
E^G	195	15	1820
E^N	0	0	0
E^I			0
Social Surplus = $E^G + E^N + E^I$	195	15	1820

These predictions motivated our treatments. In treatment **IP,** we were interested in seeing whether governments would prefer IBs to PBs, as suggested by theory, or whether the findings from the literature, that PBs are preferred to and yield higher payoffs than IBs, are robust to our calibrations and subject pool (Fehr & Schmidt 2004; Fehr, Klein & Schmidt 2007). In contrast to our theoretical prediction but in line with a related literature, PBs were

preferred to IBs and indeed outperformed IBs. Hence, in our second treatment **PS**, we compared governments' choices between the preferred PBs with SIBs.

4.3 EXPERIMENTAL IMPLEMENTATION

The experiment was programmed in Z-tree (Fischbacher, 2007), and was conducted in the Australian School of Business Experimental Lab. We recruited participants using ORSEE (Greiner 2004). We ran six sessions in November 2011 to August 2012, with additional sessions for robustness that are reported in Appendix 1. Three sessions were dedicated to the **IP** treatment and the other three to the **PS** treatment. A typical session involved 24 subjects, for a total of 138 student participants from the UNSW subject pool. Average earnings were roughly $20 an hour, as stipulated by the *ASB Lab* guidelines. Refer to Appendix 4 for more detail on demographics.

Due to the complexity of the experiment, it took on average more than an hour to go through the instructions. **PS** lasted in excess of 90 minutes. Before a session commenced, participants were given "Detailed Instructions" (Appendix 2.1) and an "Information Fact Sheet" (Appendix 2.2). Participants were given 25 minutes to read the two documents. Once the 25 minutes lapsed, participants were given 20 minutes to complete a quiz that tested their understanding (Appendix 2.3, "Experimental Quiz"). Once the 20 minutes lapsed, the experimenter (the first author in all cases) used PowerPoint to go through the questions in the quiz. Participants were encouraged to ask the experimenter questions about the instructions.

In **PS**, governments and not-for-profits had different experimental conversion currency rates to ensure that those in the role of not-for-profits would not feel deceived, in light of the recruitment announcement that promised average earnings of $30. We were also concerned about inequity concerns affecting our results; a methodological discussion of these issues may be found in Wong & Ortmann (2012b). At the beginning of the experiment, the experimenter announced that the experimental conversion currency rate was different depending on the roles they were randomly assigned to, and to ensure that governments and not-for-profits would not have hugely different expected earnings. Experimental conversion currency rates were private information (e.g. see Kagel, Kim & Moser 1996). Governments thus did not know not-for-profits' experimental currency conversion rate, nor did not-for-profits know governments' experimental currency conversion rate.

5. EXPERIMENTAL RESULTS

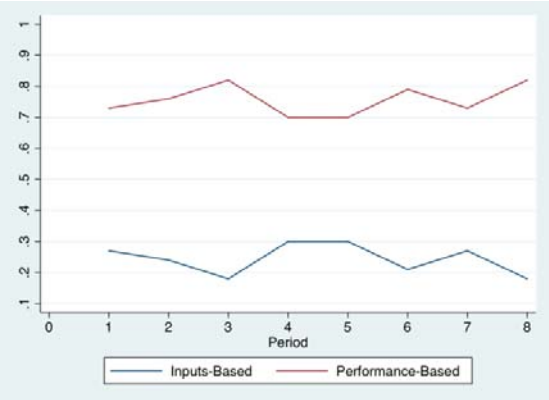
Our aim is to determine whether, and how exactly, the specific properties of each contract affect the underperformance problem plaguing not-for-profits. To this end, we study governments' behavior in each contract, including their choice of contract type offer and their designs of offers, and how these factors influence not-for-profits' effort levels. We then compare payoffs and social surplus across contracts.

5.1 TREATMENT IP

A. GOVERNMENTS' CONTRACT OFFER

In Treatment **IP**, governments can offer one of two contracts: inputs-based contracts (IBs) and performance-based contracts (PBs). If governments behave in line with theory, they will offer IBs more frequently than PBs, as it allows them to write an enforceable contract on task 1. In contrast to the theoretical predictions but in line with findings in a related work by Fehr & Schmidt (2004), governments offered PBs more frequently than IBs. Figure 5.1.1 shows the proportion of contract type offered in each period of the experiment, revealing that governments' preference for PBs was pronounced from the start of the experiment, and was never offered less than 70 percent of the time. In total, 199 (75.4 percent) of offered contracts were PBs, whereas the remaining 65 (24.6 percent) were IBs. There was no clear-cut drift.

FIGURE 5.1.1: PROPORTION OF CONTRACTS OFFERED



Result 1: Governments preferred PBs to IBs, offering them in 3 out of 4 instances.

B. GOVERNMENTS' CONTRACT OFFER DESIGN CHOICE

Table 5.1.1 contains governments' piece-rates, promised bonus b*, and actual bonus b (conditional on accepted contract offers; the relevant number of observations is mentioned in the last row), averaged over the first and second half of the experiment.

TABLE 5.1.1: SUMMARY OF GOVERNMENTS' DESIGN CHOICES

	Periods 1-4				Periods 5-8			
Contract	#	s	b*	b	#	s	b*	b
IB	33	36.1			32	40.1		
PB	99		579.6	168.9	100		853.5	243.6
				(87)				(88)

In IBs, governments must offer a piece-rate greater than or equal to the marginal cost of effort to induce payoff-maximising not-for-profits to exert effort on task 1. Recall that the marginal cost of effort is 30 when $e_1+e_2 \in [1,6]$, and increases to 75 when $e_1+e_2 \in [7,12]$. Table 5.1.1 shows that governments understood the power of the piece-rate to induce effort on task 1, as piece-rate offers were on average greater than 30.

Theory also posits that budget-maximising governments will reduce their piece-rate offer such that they just compensate not-for-profits' marginal cost of effort, but the data in Table 5.1.1 shows that average piece-rate offers rose slightly from 36.1 to 40.1 as the experiment progressed. This result is most likely due to the higher number of governments that offered piece-rates greater than or equal to 30 in the second half of the experiment than in the first, rising from 60.6 to 75.5 percent.

In PBs, budget-maximising governments are theoretically predicted to always renege on their promise to pay their bonus. We therefore predicted that governments would pay lower bonuses than what they promised. The data in Table 5.1.1 unambiguously confirms our prediction - governments' actual bonus payments were systematically and significantly lower than their promised bonus (Wilcoxon signed-rank test, $p < 0.01$). In total, governments' actual bonuses were smaller than their promised bonuses 80.5 percent of the time. This could be driven by governments that discipline underperforming not-for-profits by reducing bonus payments. This hypothesis is not confirmed, as governments underpaid their promised bonus 58.2 percent of the time, even when not-for-profits complied with their desired effort levels.

Result 2: Governments' designed contracts qualitatively in line with theoretic predictions. On average they:

a) Offered piece-rates close to the marginal cost of effort (30) in IBs;
b) Often did not pay their promised bonus in PBs, even when not-for-profits complied with their desired effort levels.

In spite of Result 2, governments' behaviour in PBs challenges the standard model of self-interest. In particular, actual bonuses increased from an average of 148.40 points in the first half of the experiment to 214.4 points in the second (Table 5.1.1). Governments also increased average bonuses from 206.5 points to 530.1 points when not-for-profits complied with their desired effort levels. Lastly, governments reciprocated high effort with high bonuses. We show this formally by replicating Fehr & Schmidt's (2004) OLS regression (1), where dependent variable *Actual Bonus* was regressed on not-for-profits' total effort $(e_1 + e_2)$, effort difference between tasks 1 and 2 $|e_1 - e_2|$, and each component of the PB contract. We also conducted a clustered OLS regression, where each cluster represents the action of a single government.[17]

$$Actual\ Bonus = ß_0 + ß_1(e_1 + e_2) + ß_2|e_1 - e_2| + ß_3 Wage + ß_4 e_1^* + ß_5 e_2^* + ß_6 b^* + e \qquad (1)$$

TABLE 5.1.2: DETERMINANTS OF ACTUAL BONUS

Dependent Variable Actual Bonus	**Robust standard Errors**	**Clusters (31 Clusters)**
Constant	-205.65***	-205.65**
	(66.75)	(92.23)
$(e_1 + e_2)$	63.94***	63.94***
	(8.12)	(13.92)
$\|e_1 - e_2\|$	-74.70***	-74.70*
	(26.51)	(38.48)
Wage	-0.51***	-0.51***
	(0.16)	(0.18)
e_1^*	11.62	11.62
	(22.12)	(23.61)
e_2^*	24.34	24.34
	(21.80)	(27.68)
b^*	-0.14	-0.14
	(0.08)	(0.08)

[17] We ran a clustered OLS regression since each observation in the normal OLS regression may not be truly independent of each other. Specifically, governments' past experience with not-for-profits might have influenced their present decisions with not-for-profits, since each experimental session ran for at least eight rounds.

Obs	175	175
Adjusted R^2	0.50	0.50

*Significance level = 5%; **Significance level = 1%; ***Significance level = 0.1%

Table 5.1.2 explains governments' actual bonus choices, conditional that not-for-profits accept the PB contract. It shows that a one-unit increase in total effort $(e_1 + e_2)$ led to a 64 point increase in *Actual Bonus* ($p < 0.01$), which suggests governments reciprocated high effort levels with high bonuses. Effort difference $|e_1 - e_2|$ had a significant ($p < 0.05$) and negative effect on *Actual Bonus*, which implies governments punished not-for-profits by reducing *Actual Bonus* by 75 points when effort difference increased by one unit. Also, governments that offered high wages paid smaller bonuses ($p < 0.01$). However its impact on *Actual Bonus* was minimal.

Result 3: Governments rewarded higher effort levels with larger bonuses.

C. NOT-FOR-PROFITS' EFFORT

The analysis thus far suggests that most governments exploited the piece-rate component in IBs and failed to deliver their promised bonus in PBs, which is in line with theoretical predictions. Governments in PBs, however, often reciprocated high effort levels with generous bonuses. This prompts the question: how did governments' choice and design of contract influence not-for-profits' performance? Table 5.1.3 summarises governments' desired effort, not-for-profits' actual effort, and the percentage of contracts that were rejected (R%) in both contracts, averaged over the first and the second half of the treatment.

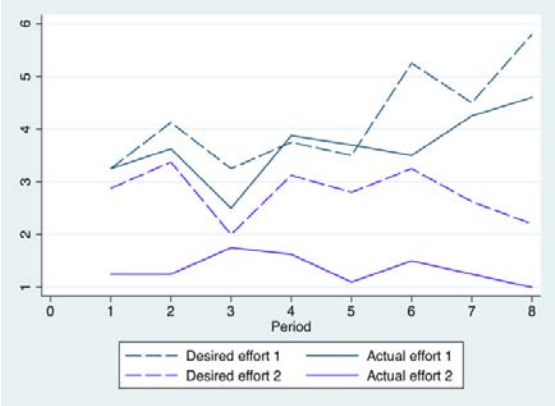
TABLE 5.1.3: SUMMARY STATISTICS OF NOT-FOR-PROFITS' BEHAVIOUR

		Periods 1-4					Periods 5-8					
Contract	#	e_1*	e_1	e_2*	e_2	R%	#	e_1*	e_1	e_2*	e_2	R%
IB	28	3.7	3.4	3.0	1.4	15	27	4.5	4.0	2.7	1.2	16
PB	87	4.3	2.7	4.2	2.6	12	88	4.5	3.1	4.6	3.0	12

In IB, when principals write a contract on one of several tasks, opportunistic agents will focus their effort on the contractible task and neglect the others. Figure 5.1.2 shows that in every period of the experiment, not-for-profits on average exerted more effort on contractible task 1 than on non-contractible task 2 (Wilcoxon signed-rank test, p-value < 0.01). In particular, not-for-profits increased effort on task 1 from 3.4 to 4.0 and reduced effort on task 2 from 1.4 to 1.2 from the first to the second half of the treatment (Table 5.1.3). In total, not-for-profits

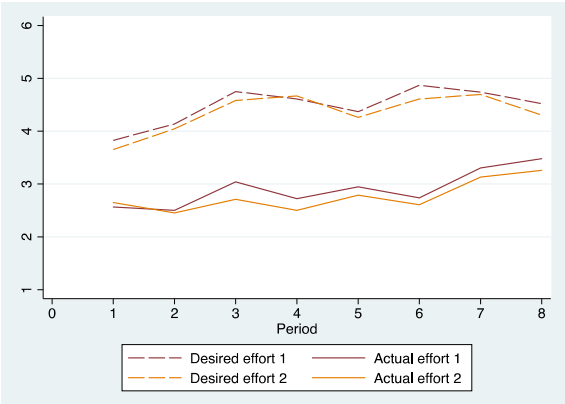
exerted lowest possible effort on non-contractible task 2 87.3 percent of the time. Also, 29.1 percent of not-for-profits exerted the theoretically predicted effort levels on tasks 1 and 2, which was also the mode effort choice.

FIGURE 5.1.2: DESIRED AND ACTUAL EFFORT IN IBs



In PBs, the non-contractible nature of task 2 allows governments to renege on their promised bonus without fear of being disciplined by a court of law. Selfish governments will thus pay an actual bonus of zero. If not-for-profits anticipate such opportunism, it is in their best interest to underperform by exerting minimum effort on tasks 1 and 2. Underperformance was prevalent, as not-for-profits systematically and significantly exerted less effort than what governments desired, exerting on average a total of 3 units less effort than what governments desired (Mann-Whitney, p-value < 0.01). This is evident in Figure 5.1.3, which compares governments' desired effort on tasks 1 and 2 to not-for-profits' actual effort on tasks 1 and 2, averaged over each period of the experiment. Also, 38.3 percent of not-for-profits exerted minimum effort on task 1 and 12.6 percent on task 2, which was also the mode effort choice in the PB.

FIGURE 5.1.3: DESIRED AND ACTUAL EFFORT IN PBs



Result 4: Not-for-profits behaved qualitatively in line with theoretic predictions. On average:

a) Not-for-profits exerted more effort on contractible task 1 than on non-contractible task 2 in IBs.
b) Not-for-profits exerted less effort than what governments desired in PBs.

In spite of Result 4, many not-for-profits exhibited behaviour in PBs that challenges the assumption of payoff-maximisation. Namely, not-for-profits exerted on average a total of 5.7 units of effort on tasks 1 and 2, even though theory predicts they will exert only 2. Moreover, not-for-profits average total effort increased from 5.3 units in the first half of the treatment to 6.3 in the second. To test the effect of bonus payments on effort levels, we conducted OLS regression (3). In the regression, dependent variable total effort $(e_1 + e_2)$ was regressed on the specific components of PBs and dummy variable $Generous^{-1}$. $Generous^{-1}$ equals 1 if, in the previous PB offer, the actual bonus payment was equal to or higher than the corresponding promised bonus, and 0 if otherwise:

$$(e_1 + e_2) = ß_0 + ß_1 e_1^* + ß_2 e_2^* + ß_3 Wage + ß_4 b^* + ß_5 Generous^{-1} + e \quad (3)$$

TABLE 5.1.4: DETERMINANTS OF TOTAL EFFORT IN PBs

Dependent Variable $(e_1 + e_2)$	**Robust Standard Errors**	**Cluster (33 Clusters)**
Constant	1.84	1.84
	(1.11)	(1.40)
e_1^*	0.43	0.43
	(0.49)	(0.53)
e_2^*	0.02	0.02
	(0.47)	(0.44)
Wage	-0.00	-0.00
	0.00	0.00
b^*	0.00*	0.00
	(0.00)	(0.00)
$Generous^{-1}$	3.01***	3.01***
	(0.60)	(0.70)
Obs	142	142
Adjusted R^2	0.26	0.26

* Significance level = 5%; ** Significance level = 1%; *** Significance level = 0.1%

Table 5.1.4 contains the regression results. $Generous^{-1}$ is highly statistically (p-value < 0.01) and economically significant in determining not-for-profits' total effort, as not-for-profits increased total effort by approximately 3 units when, in the previous PB, governments paid a

bonus that was equal to or greater than their promised bonus. The result suggests that even though the promise of a bonus is non-binding, it can be a powerful motivational tool, especially when there is some form of precedent.

Result 5: In PBs, not-for-profits exhibited behaviour that challenges the standard assumption of payoff-maximisation, as they increased total effort by approximately 3 units when they encountered a generous government in the previous PB.

D. PAYOFFS AND SOCIAL SURPLUS

Table 5.1.5 contains governments' and not-for-profits' payoffs, and social surplus (W) from the two contracts. Social surplus is the sum of all player's utility. The payoffs are in terms of experimental currency units (E), and are averaged over the first and second half of the treatment.

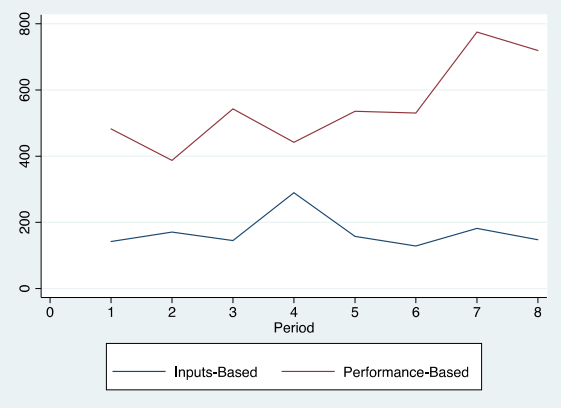
TABLE 5.1.5: SUMMARY OF PAYOFFS AND SOCIAL SURPLUS

	Periods 1-4				Periods 5-8			
Contract	#	E^G	E^N	W	#	E^G	E^N	W
IB	33	102.2	91.9	194.1	32	57.7	98.4	156.1
PB	99	350.0	116.3	466.3	100	538.5	102.8	641.3

On average, not-for-profits earned more than governments in IBs, but governments earned more than not-for-profits in PBs. However the average earnings for governments and not-for-profits were higher in PBs than in IBs. Thus, in contrast to theoretical prediction, both players were better off, and increasingly so, in PBs than in IBs.

Social surplus was unambiguously higher in PBs than in IBs, which is not surprising given effort levels were marginally higher and more efficiently allocated in PBs. This is shown in Figure 5.1.4, which shows that on average, PBs yielded higher social surplus than IBs in every period of the treatment (Mann-Whitney, p-value < 0.01). In total, average social surplus decreased in IBs from 194.1 to 156.1 points (but note the spike in period 4) but increased in PBs from 466.30 to 641.25 points from the first half to the second half of the treatment.

FIGURE 5.1.4 SOCIAL SURPLUS



Result 6: In stark contrast to theoretic predictions, PBs outperformed IBs as:

a) Not-for-profits and governments earned on average more in PBs than in IBs.
b) Average social surplus was higher in PBs than IBs in every period of the treatment. As the experiment progressed this trend became stronger.

The results from treatment **IP** show that current not-for-profit contracts fail in serious ways – governments often do not pay their promised bonus in PBs, and not-for-profits often exert low effort in PBs and disproportionate effort on task 1 in IBs, which qualitatively aligns with theoretic predictions. The results in **IP** nevertheless present an interesting puzzle: consistent with the results in Fehr and Schmidt (2004) but in contrast to theoretic predictions, PBs outperformed IBs in our setting. This poses the question: why does a contract that offers no enforceability outperform a contract with partial enforceability? The results suggest that reciprocity drove the success of PBs. But if reciprocity, and particularly the government's power to control not-for-profits' payoffs through the bonus, outperformed the mechanism of partial enforceability in treatment **IP**, is it powerful enough to outperform the mechanism of full enforceability in SIBs? Treatment **PS** answers this question.

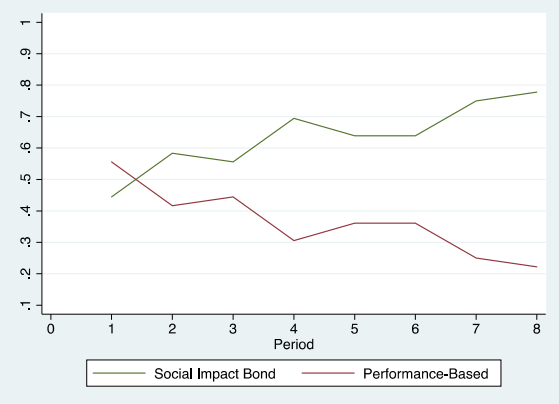
5.2 TREATMENT PS

In treatment **PS**, governments offer either performance-based contracts (PBs) or social impact bonds (SIBs) to not-for-profits.

A. GOVERNMENTS' CONTRACT OFFER

In line with theory, governments offered SIBs more often than PBs in every period but period 1 of the treatment. Figure 5.2.1 shows a steady drift towards SIBs and away from PBs, where the eventual outcome, which coincides with the theoretical prediction, seems undisputable. In total, governments offered SIBs 183 times (63.5 percent) and PBs 105 times (36.5 percent). Note that if governments' preferences are rational, by strict transitivity they would prefer SIBs to IBs, since they preferred PBs to IBs in treatment **IP**.

FIGURE 5.2.1: PROPORTION OF CONTRACTS OFFERS



Result 7: Governments offered more SIBs than PBs, and offered them 63.5 percent of the time. Contrary to **IP,** there is a clear-cut drift towards the theoretic prediction in treatment **PS**.

B. GOVERNMENTS' CONTRACT OFFER DESIGN CHOICE

Table 5.2.1 summarizes governments' design of PBs and SIBs, averaged over the first and second half of the treatment. It contains their promised bonuses b*, actual bonuses b (conditional on accepted offers; the relevant number of observations is mentioned in the second row), bond offers B, and investors' reservation bonds B^O.

TABLE 5.2.1: SUMMARY OF GOVERNMENTS' DESIGN CHOICES

	Periods 1-4					Periods 5-8				
Contract	#	b*	b	B	B^O	#	b*	b	B	B^O
PB	62	453.9	145.3			43	545.6	121.6		
			(54)					(38)		
SIB	82			701.8	688.7	101			839.2	801.0

In PBs, governments often reneged on their promise to pay the bonus. The data in Table 5.2.1 shows that governments' actual bonuses were on average at least three times lower than their promised bonuses (Wilcoxon signed-rank test, p-value < 0.01). They also underpaid their promised bonus by 68.0 percent in the first half of the treatment, which increased to 77.7 percent in the second half. In total, governments reneged on their promised bonus about 70 percent of the time (130 out of 183 PB offers. Even when not-for-profits exerted governments' desired effort levels on tasks 1 and 2, governments did not to deliver their promised bonus 53.5 percent of the time. These findings are broadly consistent with the corresponding results in treatment **IP**, where governments reneged on their promised bonus about 80 percent of the time, and reneged on their promised bonus 58.2 percent of the time even when not-for-profits complied with their desired effort levels.

Also consistent with the results in treatment **IP**, governments nevertheless exhibited behaviour in PBs that challenges the assumption of payoff-maximisation. They rewarded not-for-profits who exerted their desired effort levels on tasks 1 and 2, by increasing their average bonus payments from 135.5 to 194.8 points. To quantify the effect of total effort on actual bonus payments, we repeated OLS regression (1) from section 5.1. Table 5.2.2 contains the regression results, and shows that a one-unit increase in total effort $(e_1 + e_2)$ led to a 26 point increase in *Actual Bonus* ($p < 0.05$). Governments thus rewarded higher effort levels with larger bonus payments in both treatments.

TABLE 5.2.2: DETERMINANTS OF ACTUAL BONUS

Dependent Variable	**Robust standard errors**		**Clusters**	
Actual Bonus	**IP**	**PS**	**IP (31 Clusters)**	**PS (27 Clusters)**
Constant	-205.65***	-71.38	-205.65**	-71.38
	(66.75)	(74.43)	(92.23)	(86.36)
$(e_1 + e_2)$	63.94***	25.63**	63.94***	25.63***
	(8.12)	(10.13)	(13.92)	(9.14)
$\lvert e_1 - e_2 \rvert$	-74.70***	-5.86	-74.70*	-5.86
	(26.51)	(33.71)	(38.48)	(38.97)
Wage	-0.51***	0.19	-0.51***	0.19
	(0.16)	(0.12)	(0.18)	(0.12)
e_1*	11.62	-16.52	11.62	-16.52
	(22.12)	(25.44)	(23.61)	(23.42)
e_2*	24.34	-9.96	24.34	-9.96
	(21.80)	(18.11)	(27.68)	(14.04)
*b**	-0.14	0.24**	-0.14	0.24*
	(0.08)	(0.10)	(0.08)	(0.12)
Obs	175	92	175	92
Adjusted R^2	0.50	0.28	0.50	0.28

* Significance level = 5%; ** Significance level = 1%; *** Significance level = 0.1%

In SIBs, the presence of an investor who can obtain hard monitoring information about not-for-profits' performance allows the investor (and thus governments) to perfectly enforce effort on tasks 1 and 2. Governments should thus desire the highest possible effort on tasks 1 and 2. Moreover, the presence of a rational, computerized investor allows governments to fully extract investors' surplus, since other factors such as social preferences or human error cannot interfere with investors' decision-making. We thus ask: did governments fully exploit investors' ability to obtain hard monitoring information and rationality?

The data in Table 5.2.1 shows that governments did not manage to fully exploit the perfect enforceability of SIBs, as they desired the theoretically predicted effort levels only 53.6 percent of the time (98 out of 183 SIB offers). On average they desired a total of 9.6 units of effort, which is lower than the theoretically predicted 12, and made a bond offer of 777.6 points, which is lower than the theoretically predicted 880. Averages, however, are misleading in the present context. Governments asked for higher effort levels as the treatment progressed and their average bond offer and investors' reservation bond increased from 701.8 and 688.7 points respectively in the first half of the treatment, to 839.2 and 801.0 points respectively in the second. (Recall that the theoretically predicted bond offer and reservation bond is 880 points.) Governments also often exploited investors' rationality, as 82.4 percent

of accepted bond offers were very close to the investors' reservation bond (i.e. reservation bond ≤ actual bond offer < reservation bond + 50). There were only 7 times where the investor rejected SIB (i.e. bond < minimum bond). It therefore appears that governments (increasingly) behaved in line with theoretical predictions.

Result 8: Governments' designed contracts that were qualitatively in line with theoretic predictions as:

a) In PBs, governments often did not pay their promised bonus, even when not-for-profits complied with their desired effort levels.
b) In SIBs, governments offered bonds that were close to the investor's reservation bond, and demanded higher effort levels as the experiment progressed.

Result 9: In spite of Result 8, some governments exhibited behaviour that deviated from theoretic predictions. Namely,

a) Governments sometimes reciprocated high effort with generous bonuses in PBs.
b) Governments desired highest effort levels $(e_1^*, e_2^*) = (6,6)$ only 53.6 percent of the time, although much of this behavior might be due to learning as there is clear-cut drift towards the theoretic prediction as the treatment progressed.

C. NOT-FOR-PROFITS' EFFORT

Table 5.2.3 contains governments' desired effort levels, not-for-profits' actual effort levels, and not-for-profits' rejection rate (R%), averaged over the first and second half of the treatment.

TABLE 5.2.3: SUMMARY STATISTICS OF NOT-FOR-PROFITS' BEHAVIOUR

		Periods 1-4					Periods 5-8					
Contract	#[18]	e_1^*	e_1	e_2^*	e_2	R%	#	e_1^*	e_1	e_2^*	e_2	R%
PB	54	3.9	2.9	3.7	2.9	13	38	4.3	2.4	4.5	2.5	12
SIB	76	4.8	4.8	4.8	4.8	7	89	5.4	5.3	5.5	5.3	12

Not-for-profits qualitatively behaved in line with theoretic predictions, as they on average exerted less than what governments desired in PBs (Mann-Whitney, $p < 0.01$; see Figure 5.2.2), but often complied with governments' desired effort levels in SIBs[19].

[18] Observations for desired effort levels (because the number of observations for desired and actual effort levels would obviously be different due to nfp service providers rejecting contracts).

FIGURE 5.2.2: DESIRED AND ACTUAL EFFORT LEVELS IN PBs

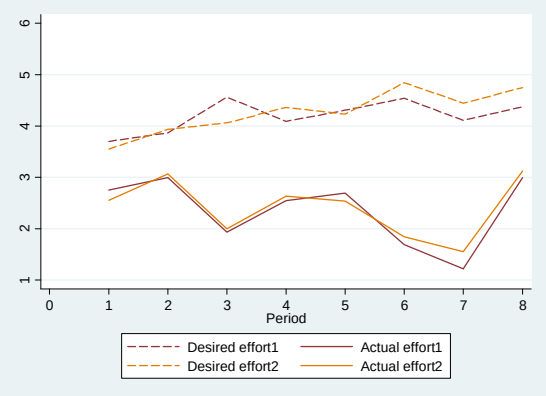
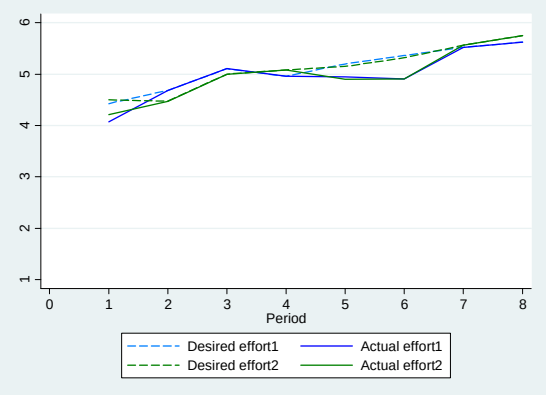


FIGURE 5.2.3: DESIRED AND ACTUAL EFFORT LEVELS IN SIBs



Result 10: Not-for-profits behaved qualitatively in line with theoretic predictions, as they on average:

a) Exerted less effort than what governments desired in PBs.

b) Complied with governments' desired effort levels in SIBs.

[19] There were four times where not-for-profits did not comply with governments' desired effort levels: 1) Session 1, Period 1: The government specified (e_1*, e_2*) = (3, 4), but the not-for-profit exerted (e_1, e_2) = (3, 5). This might be due to a "trembling hand"; 2) Session 2, Period 1: The government specified (e_1*, e_2*) = (6, 6), but the not-for-profit exerted (e_1, e_2) = (1, 1). This was most likely a mistake. In all subsequent periods, the subject complied with governments' desired effort levels in SIBs; 3) Session 2, Period 5: The government specified (e_1*, e_2*) = (6, 6), but the not-for-profit exerted (e_1, e_2) = (1, 1). The subject complied with government's desired effort levels previously, but rejected every subsequent SIB offer where (e_1*, e_2*) = (6, 6); 4) Session 2, Period 6: The government specified (e_1*, e_2*) = (6, 6), but the not-for-profit exerted (e_1, e_2) = (1, 1). This was the first time the subject got a SIB offer, so it could be a mistake.

TABLE 5.2.4: DISAGGREGATED EFFORT LEVELS

i) PB

		Effort 2					
		1	**2**	**3**	**4**	**5**	**6**
Effort 1	**1**	33	1	0	0	1	0
	2	5	1	1	1	0	0
	3	0	2	15	4	1	0
	4	0	1	2	5	2	0
	5	0	0	0	2	6	0
	6	0	0	0	0	0	9

ii) SIB

		Effort 2					
		1	**2**	**3**	**4**	**5**	**6**
Effort 1	**1**	5	0	0	0	0	0
	2	0	0	1	0	1	0
	3	0	2	9	7	1	0
	4	0	2	4	11	6	0
	5	0	0	1	5	9	8
	6	0	0	0	1	6	86

Table 5.2.4 illustrates the distribution of efforts for the two contracts. For accepted PB offers, there are spikes of decreasing prominence at $(e_1, e_2) = (1,1)$, (3,3), and (6,6), with the mode coinciding with the theoretical prediction. For accepted SIB offers, there is one prominent spike at $(e_1, e_2) = (6,6)$, which coincides with the theoretical prediction. Not-for-profits' effort levels in PBs thus mirror their effort levels in SIBs.

At first glance, these results may seem unsurprising in light of the investor's ability to perfectly enforce effort on tasks 1 and 2 in SIBs, and governments' inability to enforce effort in PBs. However recall that theory predicts that all not-for-profits will exert minimum effort in PBs, but not-for-profits on average exerted a total of 5.5 units of effort. In line with treatment **IP**, not-for-profits deviated from this prediction, as they increased total effort by approximately 1.8 units when, in the previous PB, governments paid a bonus that was equal to or greater than their promised bonus (refer to Table 5.2.5).

TABLE 5.2.5: OLS ON TOTAL EFFORT IN PBs

Dependent Variable ($e_1 + e_2$)	**Robust Standard Errors**	**Cluster (29 Clusters)**
Constant	2.61	2.61
	(1.59)	(1.78)
e_1*	0.29	0.29
	(0.37)	(0.38)
e_2*	0.22	0.22
	(0.40)	(0.38)
Wage	-0.00	-0.00
	0.00	0.00
*b**	-0.00	-0.00
	(0.00)	(0.00)
$Generous^{-1}$	1.84**	1.84*
	(0.86)	(0.95)
Obs	58	58
Adjusted R^2	0.10	0.10

* Significance level = 5%; ** Significance level = 1%; *** Significance level = 0.1%

Result 11: In spite of Result 8, some governments exhibited behaviour that deviated from theoretic predictions. Namely, they

a) Exerted more than what was theoretically predicted in PBs.
b) Increased total effort by approximately 1.8 units when they encountered a generous government in the previous PB.

Lastly, not-for-profits rejected 18 out of the 183 SIB offers. Though theory predicts not-for-profits will never reject SIB offers since receiving a wage is better than nothing, the rejection rate is lower than what is commonly observed in the literature, especially since the equivalent of governments often earn substantially more experimental points than not-for-profits (e.g., see Camerer, 2003; Fehr & Schmidt, 2004; but see again our previous caveat about the publicly announced difference in ECU conversion rates).

D. PAYOFFS AND SOCIAL SURPLUS

To determine which contract fared best, we compare payoffs and social surplus in PBs and SIBs. Table 5.2.6 contains governments' budget, not-for-profits' utility, the investor's utility, and social surplus in both contracts in experimental points, averaged over the first and second half of the treatment.

TABLE 5.2.6: PAYOFFS AND SOCIAL SURPLUS

Contract	Periods 1-4					Periods 5-8				
	#	M^G	M^N	M^I	W	#	M^G	M^N	M^I	W
PB	62	285.5	233.4		519.0	43	94.9	338.0		432.9
SIB	82	1031.8	175.7	18.9	1226.4	101	1220.9	168.5	35.1	1424.5

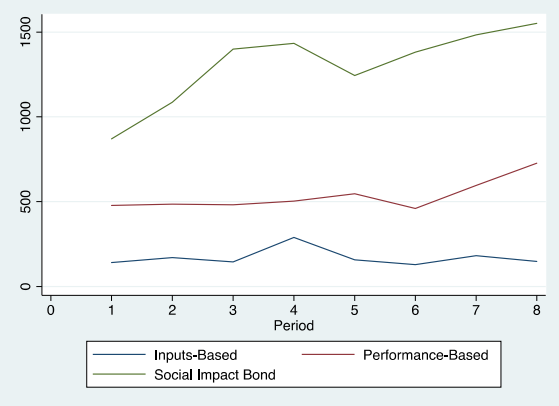
Table 5.2.6 shows that governments earned substantially higher payoffs in SIBs than not-for-profits, earning on average at least five times more points than not-for-profits. In contrast, not-for-profits earned on average higher payoffs in PBs than governments, which is inconsistent with the results obtained in **IP**[20]. However average wage offers were 321.0 points in the first half of the treatment, which increased to 430.1 points in the second, and not-for-profits' average cost of effort was 198.4 points in the first half of the treatment, which decreased to 179.3 points in the second. It is therefore possible that the earnings in PBs is driven by governments who attempted to elicit gift exchange from not-for-profits by offering high wages (Fehr, Kirchsteiger, & Riedl, 1993), but got burnt by underperforming not-for-profits.

Social surplus was higher in SIBs than in PBs, as the perfect enforceability offered by investors led to higher and more efficient effort levels in SIBs[21]. On average, social surplus in PBs was 483.7 points, and 1335.7 points in SIBs. SIBs thus unequivocally outperformed PBs on the basis of social surplus. If behaviour is roughly consistent across treatments, it implies that SIBs will outperform IBs, since PBs outperformed IBs in Treatment **IP**. Indeed Figure 5.2.4 compares the social surplus generated in each contract from Treatments **IP** and **PS**, and shows that SIBs unambiguously yielded the highest social surplus, followed by PBs and IPs.

[20] The results from Treatment **IP** may because inconsistent to that of Treatment **PS,** due to the higher proportion of local students in Treatment **PS**.

[21] The results differ from those in Fehr, Klein and Schmidt (2007); they find that bonus contracts, which are akin to PBs, are offered more and outperform incentives contracts, which are akin to SIBs (see Section 6.2 for further discussion).

FIGURE 5.2.4: SOCIAL SURPLUS POOLED DATA



Result 12: SIBs unambiguously generated higher surplus than PBs, and thus outperformed PBs. If behaviour is consistent across treatments, the result implies that SIBs would also outperform IBs.

6. DISCUSSION

6.1. DO SOCIAL IMPACT BONDS WORK? AND DOES THEIR PERFORMANCE COMPARE TO THAT OF THEIR COMPETITORS?

Our aim was to explore whether, and how exactly, the incentives embedded in three prominent contracts exacerbate (or allay) the underperformance problem afflicting not-for-profits. The results from our treatment **IP** show that governments' inability to monitor not-for-profits exacerbates the underperformance problem (and hence is likely to address the underfunding problem that many not-for-profits seem to be afflicted with). In particular, the piece-rate component in IBs induced not-for-profits to exert disproportionate effort on task 1 and neglect effort on task 2. The non-binding nature of the bonus in PBs induced governments to renege on their promise to pay the bonus, which in turn caused not-for-profits to exert low effort levels. In spite of this, in the absence of an investor who can obtain hard monitoring information about not-for-profits' performance, governments preferred PBs to IBs, as they induced higher effort than IBs. Namely, a fair number of governments paid generous bonuses to reward hard working not-for-profits, which induced a fair number of not-for-profits to exert high effort in PBs. As a result of this reciprocity "mechanism", PBs outperformed IBs. Our results are consistent with some of the literature, in spite of our different calibration.

In light of this result, in treatment **PS** we determined how SIBs fare in comparison to PBs. We thus transferred two assumptions from the models developed in Section 3 – that investors are rational and able to obtain hard monitoring information about not-for-profits' performance – to the laboratory to test whether SIBs work when they possess the ingredients that induce first-best effort levels theoretically.

The predictions of treatment **PS** are clear – SIBs should outperform PBs since they offer perfect enforceability. However perfect enforceability comes at a cost – governments must relinquish control over not-for-profits' payoffs to the investor. The theoretic and experimental literature often portrays the investor or "middleman" as the archetypal selfish economic actor (Mookherjee 2006; Plott 1986). The middleman's sole objective is to profit-maximize, and so extracts maximum surplus from agents each time they transact. To preserve this assumption and to truly highlight the consequence of delegation, we used a computerized investor who was programmed to maximize payoffs in the experiment. The investor pays the not-for-profit a wage of $200 + c(e_1^*, e_2^*)$ when he exerts the government's desired effort on tasks 1 and 2, which yields him a net payoff of 200 points (=$2). Hence, our rational investor never reciprocates high effort with high wages. By using the perfect enforceability of SIBs, governments lose control over not-for-profits' payoffs and the ability to motivate them through reciprocity, and so the mechanism that rendered PBs powerful is absent in SIBs.

In spite of this drawback, the results from treatment **PS** suggest SIBs work well: they were preferred to and unambiguously yielded higher social surplus than PBs. Moreover, not-for-profits' effort levels converged towards the social optimum effort levels as the treatment progressed. Our experiment thus shows that SIBs work in a first-best world. Investors' ability to perfectly enforce effort can overcome the drawback associated with governments' loss of control and their inability to reciprocate good performance with generous wages.

We readily admit that SIBs in the real world outside of the laboratory are highly unlikely to be implemented in a first-best environment, and that our results are driven by transferring two assumptions from the models developed in Section 3 to treatment **PS** – that investors are rational and able to obtain hard monitoring information about not-for-profits' performance. Obviously, those assumptions – in our view in particular the second one – need to be relaxed in future work.

6.2 BEHAVIOURAL IMPLICATIONS

We observed behaviour that substantially deviated from the standard assumption of payoff-maximisation. We thus take a detour from our main research question and attempt to explain these anomalies.

Models of self-regarding preferences cannot explain why PBs were preferred to, and induced higher effort than IBs. Theory predicts that selfish governments will never pay the voluntary bonus, and so not-for-profits will always exert the lowest possible effort on tasks 1 and 2. Governments should therefore offer IBs, since they offer partial enforceability through the piece-rate. In the experiment, the promise of a bonus was a more powerful motivational tool than the piece-rate, causing PBs to outperform IBs in every possible dimension: they were preferred by governments, induced higher and more efficient effort levels, and yielded higher payoffs and social surplus than IBs.

Why then was the promise of a bonus a better motivational tool than the partial enforceability of the piece-rate? In our experiment, we used complete anonymity and random matching to remove opportunities for reputation building and to decrease the probability of future interactions. We therefore attempted to remove important explicit incentives to behave prosocially (Burnham & Johnson 2005), though it is possible that our lack of asset legitimacy and social distance contributed to our results (Cherry et al, 2002).

Indeed there is a growing body of literature that questions the assumption of self-interested man, claiming that experimental, anthropological and historical evidence lends support to the existence of homo reciprocans – an individual who behaves prosocially even at potentially a personal cost. His actions are often motivated by concerns for fairness, inequity aversion or reciprocity (Charness & Rabin 2002; Fehr & Schmidt 1999; Bolton & Ockenfels 2000; Rabin 1993). His other-regarding preferences can, even if he is just one of several types making an appearance in a particular situation, alter the incentive properties of some contracts (e.g., see Fehr, Kirchsteiger & Riedl 1993; Fehr & Schmidt 2004; Fehr, Klein, & Schmidt 2007). In our experiment, it appears that some governments in PBs behaved in line with homo reciprocans. Namely, they were willing to pay generous bonuses to reward high performing not-for-profits, even though it was individually costly.[22] As such, not-for-profits exerted higher effort levels.[23] Whether that was because they were reciprocal or wanted to appear reciprocal cannot

[22]Though it is beyond the scope of this paper to identify the exact cause of this reciprocity, Fehr & Schmidt (2004) and Fehr, Klein, & Schmidt (2007) attribute it to inequity aversion. For a detailed reading of inequity aversion, refer to Fehr & Schmidt (1999).

[23] Note that governments could also use reciprocity in IBs. Specifically, they could offer large wages to elicit high effort levels or "gift exchange" from not-for-profits (e.g. see Fehr, Kirchsteiger, & Riedl 1993; List 2006; Gneezy & List 2006). However such a contract would be tremendously risky for

be determined through our study, but is an interesting methodological issue (e.g., Dana et al. 2007; see also Zizzo 2010).

Our experimental results pose a potentially puzzling question – why does the reciprocity "mechanism" in PBs outperform the partial enforceability of the piece-rate in IBs, but the full enforceability of the investor in SIBs outperform the reciprocity mechanism in PBs? In other words, why do governments display other-regarding preferences in Treatment **IP** but *less* so in Treatment **PS**? And why do not-for-profits accept SIB offers so readily? Indeed, at first sight it seems surprising that not-for-profits rarely rejected SIB offers in spite of the seemingly considerable payoff inequality that often resulted (see, for example, Fehr & Schmidt 1999, 2006; Camerer 2003). Namely, if governments specified a performance target $\{e_1, e_2\} = \{6, 6\}$, they would earn 1820 points (=$9) whereas not-for-profits would only earn 200 points (=$2).[24]

A key caveat for experimental economics is that the results of any experiment are contingent on the specifics of its design and implementation (Camerer 2003; Smith 2002; Ortmann 2010). Hence, perhaps the design of our experiment – which uses a rational and computerized investor and gives governments the choice of one of two contracts – may have driven governments away from the reciprocity "mechanism" in PBs to the safe but mechanical mechanism in SIBs. Specifically, governments offer either PBs or SIBs to not-for-profits in Treatment **PS**. Not-for-profits can burn governments in PBs by exerting low effort, whereas governments are protected from being burned in SIBs by virtue of the investor. By making this shift in power explicit, governments might justify making SIBs offers and not-for-profits might legitimize them – even though it means governments lose control over not-for-profits' payoff and their ability to exercise reciprocity via the bonus (Schnedler & Vadovic, 2011). SIB offers may further be legitimized when governments "give up" earning more by desiring less than socially optimum effort levels. Therefore the specific design of our experiment may have, in part, driven the success of SIBs.

---

governments, since there is some probability they might encounter self-interested not-for-profits. This explains why PBs – where governments can reciprocate high effort with generous bonuses rather than eliciting reciprocity through paying large upfront wages – were most probably preferred to IBs.

[24] We remind the reader that we told subjects ex ante that the exchange ratios were different for governments and service providers and were calibrated such that expected earnings were about the same regardless of the role that subjects were assigned. We therefore doubt, also in light of the considerable drift towards the theoretic prediction in our treatment IP, that it is really reciprocity that drives our results (see Dana et al. 2007; Kagel et al. 1996)

7. CONCLUDING REMARKS AND FUTURE RESEARCH

Policy-makers world-wide set high hopes in social impact bonds. Given the high hopes governments on various levels in England, Australia, and New York, for example, have pinned on this contract format, the considerable amount of money that is currently poured into this emerging market, and the fact that serious are program evaluations cannot be expected any time soon (Disley et al. 2011; see also McKay 2013 and Pratt 2013), we have provided a first test of this new contract by way of experimental methods. Specifically, we report an investigation of how SIBs perform in a first-best world, where investors are rational and able to obtain hard information about not-for-profits' performance. To this end, we used a principal-agent multi-tasking framework to compare SIBs to inputs-based (IBs) and performance-based (PBs) contracts, which represent the most commonly used contracts governments and not-for-profits write.

We have shown both theoretically and experimentally how current contracting mechanisms exacerbate the underperformance problem afflicting not-for-profits, and how social impact bonds can overcome the underperformance problem in a first-best world. Our results therefore suggest that, at least for our experimental test-bed and the embedded assumptions of rationality and availability of hard information, social impact bonds can allay the underperformance problem afflicting not-for-profits in a first-best world. This statement, of course, does not say anything about what might happen in different worlds. By testing social impact bonds in a first-best environment, we have provided a platform which other theories and/or experiments can take as point of departure. In particular, we can now investigate how SIBs will fare both theoretically and experimentally relative to other contracts when the assumptions that make them first-best are relaxed.

REFERENCES

ARTD Consultants. (2007). *Evaluation Report: NSW Youth Partnerships with Pacific Communities.* Sydney: ARTD Consultants. Retrieved from: http://www.sydneyr.det.nsw.edu.au/equity/documents/TSandR/FinalReportYPPCEvaluationARTD310707.pdf

Attorney General & Justice. (2011). *2010/11 Annual Report. A just and safe society.* Parramatta: Attorney General & Justice. Retrieved from: http://www.correctiveservices.nsw.gov.au/__data/assets/pdf_file/0011/336359/ag-and-j-2010-11-annual-report.pdf

Auditor-General New South Wales. (2010). *Auditor-General's Report Financial Audits Volume Six 2010.* Sydney: Auditor-General New South Wales.

Bacchiega, A., & Borzaga, C. (2003). The Economics of the Third Sector: Toward a More Comprehensive Approach. In H. Anheier, & A. Ben-Ner, *The study of nonprofit enterprise: Theories and approaches* (pp. 27-48). Ann Arbor, MI: University of Michigan Press.

Badelt, C. (1997). Entrepreneurship theories of the non-profit sector. *Voluntas: International Journal of Voluntary and Nonprofit Organizations, 8* (2), 162-178.

BDO (2012). *BDO Not-For-Profit Fraud Survey 2012*. Retrieved May 30, 2012, from BDO Australia: http://www.bdo.com.au/resources/surveys/forensic/bdo-not-for-profit-fraud-survey-2012

Beattie, V., Goodacre, A., Pratt, K., & Stevenson, J. (2001). The determinants of audit fees - evidence from the voluntary sector. *Accounting and Business Research, 31* (4), 243-274.

Ben-Ner, A., & Gui, B. (2003). The Theory of Nonprofit Organizations Revisited. In H. Anheier, & A. Ben-Ner (Eds.). Ann Arbor, MI: University of Michigan Press.

Berg, J., Dickhaut, J., & McCabe, K. A. (1995). Trust, reciprocity and social history. *Games and Economic Behavior,* 10, 122-142.

Bernholz, L. (2011, 4 18). *Tea, lives, and trust*. Retrieved May 30, 2012, from Philanthropy 2173 The Future of Good: http://philanthropy.blogspot.com.au/2011/04/tea-lives-and-trust.html

Bolton, E. (2010). *Social Impact Bonds: Unlocking investment in rehabilitation.* London UK: Social Finance. Retrieved from: http://www.socialfinanceus.org/pubs/socialfinance/social-impact-bonds-unlocking-investment-rehabilitation

Bolton, E., & Savell, L. (2010). *Towards a New Social Economy: Blended Value Creation through Social Impact Bonds.* Social Finance UK, London, UK. Retrieved from: http://www.socialfinance.org.uk/sites/default/files/Towards_A_New_Social_Economy_web.pdf

Bolton, G. E., & Ockenfels, A. (2000). A Theory of Equity, Reciprocity, and Competition. *The American Economic Review , 90* (1), 166-193.

Boris, E. T., de Leon, E., Roeger, K. L., & Nikolova, M. (2010). *Human Service Nonprofits and Government Collaboration. Findings from the 2010 National Survey of Nonprofit Government Contracting and Grants.* Centre on Nonprofits and Philanthropy. Washington: Urban Institute. Retrieved from: http://www.urban.org/publications/412228.html

Burnham, T., & Johnson, D. (2005). The biological and evolutionary logic of human cooperation. *Analyse & Kritik , 27*, 113-135.

Camerer, C. F. (2003). *Behavioral game theory: experiments in strategic interactions.* Princeton: Princeton University Press.

Centre for Social Impact. (2011). *Report on the NSW Government Social Impact Bond Pilot.* Sydney.

Charness, G., & Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *The Quarterly Journal of Economics, 117* (3), 817-869.

Cunneen, C., Luke, G., & Ralph, N. (2006). *Evaluation of the Aboriginal Over-Representation Strategy.* Institute of Criminology; University of Sydney Law School. Retrieved from: http://www.djj.nsw.gov.au/pdf_htm/publications/research/AORSP_Final_Report.pdf

Cunneen, C., Luke, G., Grix, J., Boyd-Caine, T., Ralph, N., & Lugg, J. (2005). *Evaluation of the Post Release Support Program.* Sydney: NSW Department of Juvenile Justice (unpublished).

Dana, J., Weber, R., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusionary preference for fairness. *Economic Theory,* 33, 67-80.

Disley, E., Rubin, J., Scraggs, E., Burrowes, N., Culley, D., & Europe, R. (2011). *Lessons learned from the planning and early implementation of the Social Impact Bond at HMP Peterborough.* Ministry of Justice . Ministry of Justice. Retrieved from: http://www.justice.gov.uk/downloads/publications/research-and-analysis/moj-research/social-impact-bond-hmp-peterborough.pdf

Engelmann, D., & Ortmann, A. (2009). The Robustness of Laboratory Gift Exchange: A Reconsideration. *Working paper, Charles University* .

Erus, B., & Weisbrod, B. A. (2003). Objective functions and compensation structures in nonprofit and for-profit organizations: Evidence from the "Mixed" hospital industry. In E. L. Glaeser, *The governance of not-for-profit organizations* (pp. 117-142). Chicago: IL: University of Chicago Press.

Erwin, d. L. (2011). *Nonprofit-Government Contracting in the Nation's Capital: Challenges and Opportunities.* Centre on Nonprofits and Philanthropy. Washington: Urban Institute.

Fehr, E., & Schmidt, K. M. (1999). A Theory of Fairness, Competition and Co-operation. *Quarterly Journal of Economics, 114*, 817-868.

Fehr, E., & Schmidt, K. M. (2004). Fairness and Incentives in a Multi-task Principal-Agent Model. *Scandinavian Journal of Economics, 106* (3), 453-474.

Fehr, E., & Schmidt, K. M. (2006). The Economics of Fairness, Reciprocity and Altruism - Experimental Evidence and New Theories. *Handbook on the Economics of Giving, Reciprocity and Altruism* .

Fehr, E., Kirchsteiger, G., & Riedl, A. (1993). Does Fairness Prevent Market Clearing? An Experimental Investigation. *The Quarterly Journal of Economics, 108* (2), 437-59.

Fehr, E., Klein, A., & Schmidt, K. M. (2007). Fairness and Contract Design. *Econometrica, 75* (1), 121-154.

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Ready-made Economic Experiments. *Experimental Economics, 10* (2), 171-178.

Fréchette, G. R. (2012). Session-effects in the laboratory. *Experimental Economics*, *15*(3), 485-498.

Friesen, L., & Gangadharan, L. (2012a). Individual level evidence of dishonesty and the gender effect. *Economics Letters*, 117, 624-626.

Friesen, L., & Gangadharan, L. (2012b). Designing Self-Reporting Regimes to Encourage Truth Telling: An Experimental Study. *Discussion Papers Series 426, School of Economics, University of Queensland, Australia.*

Glaeser, E. L. (2003). *The governance of not-for-profit organizations.* Chicago: University of Chicago Press.

Gneezy, U., & List, J. A. (2006). Putting Behavioural Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments. *Econometrica, 74* (5), 1365-1384.

Greiner, B. (2004). The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics. *Working Paper Series in Economics 10, University of Cologne, Department of Economics.*

Guala, F. (2010). Reciprocity: weak or strong? What punishment experiments do (and do not) demonstrate. *Departmental Working Papers 2010-23, Department of Economics, Management and Quantitative Methods at Università degli Studi di Milano* .

Hansmann, H. (1987). Economic Theories of Nonprofit Organization. In W. W. Powell (Ed.), *The Nonprofit Sector A Research Handbook* (pp. 27-42). Yale University Press.

Hansmann, H. (1980). The role of nonprofit enterprise. *Yale Law Journal, 89* (5), 835-901.

Hansmann, H. (2003). The Role of Trust in Nonprofit Enterprise. In H. Anheier, & A. Ben-Ner (Eds.), *The study of nonprofit enterprise: Theories and approaches* (pp. 115-122). The study of nonprofit enterprise: Theories and approaches: The University of Michigan Press.

Helmut K. Anheier and Stefan Toepler (eds.). *International Encyclopedia of Civil Society*. New York, NY: Springer, 2010.

Holmes, J. (2011). Re-offending in NSW. *Crime and Justice Statistics - Bureau Brief* (56). Retrieved from: http://www.bocsar.nsw.gov.au/lawlink/bocsar/ll_bocsar.nsf/vwFiles/bb56.pdf/$file/bb56.pdf

Holmstrom, B., & Milgrom, P. (1991). Multi-Task Principal-Agent Problems: Incentive Contracts, Asset Ownership and Job Design. *Journal of Law, Economics and Organization, 7* (Special Issue), 24-52.

Kagel, J., Kim, C., & Moser, D. (1996). Fairness in Ultimatum Games with Asymmetric Information and Asymmetric Payoffs. *Games and Economic Behavior, 13* (1), 100-110.

Klein, B., & Leffler, K. (1981). The role of market forces in assuring contractual compliance. *Journal of Policy Economy, 81*, 615-640.

Kocher, M., & Sutter, M. (2007). Individual versus group behavior and the role of the decision making procedure in gift-exchange experiments. *Empirica, 34* (1), 63-88.

Kreps, D. M. (1990). Corporate culture and economic theory. In J. Alt, & K. Shepsle (Eds.), *Perspectives on positive political economy* (pp. 90-143). Cambridge, UK: Cambridge University Press.

Krishnan, R., Yetman, M., & Yetman, R. J. (2006). Expense Misreporting in Nonprofit Organizations. *Accounting Review, 81* (2), 399-420.

Kristof, N. D. (2011, April 11). *'Three Cups of Tea,' Spilled*. Retrieved May 30, 2012, from The New York Times: http://www.nytimes.com/2011/04/21/opinion/21kristof.html

Kugler, T., Borstein, G., Kocher, M., & Sutter, M. (2007). Trust between individuals and groups: Groups are less trusting than individuals but just as trustworthy. *Journal of Economic Psychology, 28* (6), 646-657.

Liebman, J. B. (2011). *Social Impact Bonds: A promising new financing model to accelerate social innovation and improve government performance.* Washington DC: American Progress.

List, J. A. (2006). The Behaviouralist Meets the Market: Measuring Social Preferences and Reputation Effects in Actual Transactions. *Journal of Political Economy, 114* (1), 1-37.

Luhan, W. J., Kocher, M. G., & Sutter, M. (2009). Group polarization in the team dictator game reconsidered. *Experimental Economics, 12* (1), 26-41.

McKay, K. (2013, April 8), Debunking the Myths Behind Social Impact Bond Speculation. Stanford Social Innovation Review. Retrieved from: http://www.ssireview.org/blog/entry/debunking_the_myths_behind_social_impact_bond_speculation

Mookherjee, D. (2006). Decentralization, Hierarchies, and Incentives: A Mechanism Design Perspective. *Journal of Economic Literature, 44* (2), 367-390.

Mulgan, G., Reeder, N., Aylott, M., & Bo'Sher, L. (2010a). *Social Impact Investment: The opportunity and challenge of Social Impact Bonds.* London UK: The Young Foundation. Retrieved from: http://evpa.eu.com/wp-content/uploads/2011/06/11-04-11_Social_Impact_Investment_Paper_2.pdf

Mulgan, L. J., Reeder, N., & Shelupanov, A. (2010b) *Financing social value: implementing Social Impact Bonds.* London UK: The Young Foundation. Retrieved from: http://youngfoundation.org/files/images/01_10_Socail_Impact_Bonds_FINAL.pdf

Murphy, P., McGinnes, A., Balmaks, A., McDermott, T., & Corriea, M. (2010). *A Strategic Review of the New South Wales Juvenile Justice System. Report for the Minister of Juvenile Justice.* Manuka, ACT: Noetic Solutions Pty Limited. Retrieved from: http://www.djj.nsw.gov.au/pdf_htm/publications/general/Juvenile%20Justice%20Review%20Report%20FINAL.pdf

NVCO (National Council for Voluntary Organizations, in association with Deloitte) (2011), Managing Risk. A New Approach for the Fit and Proper Regulation of Charities. Retrieved from: http://www.ncvo-vol.org.uk/sites/default/files/Task_Finish_Group_report_-_final_July_2011.pdf

NSW Bureau of Crime Statistics and Research. (2010). *New South Wales Criminal Court Statistics.* Sydney: NSW Bureau of Crime Statistics and Research. Retrieved from: http://www.bocsar.nsw.gov.au/lawlink/bocsar/ll_bocsar.nsf/vwFiles/CCS2010.pdf/$file/CCS2010.pdf

Ortmann, A. (1996). Modern economic theory and the study of nonprofit organizations: Why the Twain shall meet. *Nonprofit and Voluntary Sector Quarterly, 25*, 470-484.

Ortmann, A. (2010). 'The Way in Which an Experiment is Conducted in Unbelievably Important:': On the Experimentation Practices of Eonomists and Psychologists. ", pp. 38 - 55 in E.H. Witte & T. Gollan (eds.), Tagungsband zum 25. Hamburger Symposion - Schwerpunktthema: Sozialpsychologie und Ökonomie, Pabst Verlag 2010.

Ortmann, A., & Hertwig, R. (2002). The Costs of Deception: Evidence from Psychology. *Experimental Economics, 5* (2), 111-131.

Ortmann, A., & Schlesinger, M. (2003). Trust, repute, and the role of nonprofit enterprise. In H. Anheier, & A. Ben-Ner, *The study of nonprofit enterprise: Theories and approaches* (pp. 77-114). Ann Arbor, MI: University of Michigan Press.

Plott, C. (1987). Dimensions of Parallelism: Some Policy Applications of Experimental Methods. In A. E. Roth (Ed.), *Laboratory Experimentation in Economics: Six Points of View.* New York: Cambridge University Press.

Plott, C. R. (1986). Rational Choice in Experimental Markets. *The Journal of Business, 59* (4 Part 2), 301-327.

Powell, W.W., & Steinberg, R. (2006), The Nonprofit Sector: A Research Handbook, Second Edition. Yale: Yale University Press

Pratt, J. (2013, April 17), Flaws in the Social Impact Bond/Pay for Success Craze. Retrieved from: http://www.nonprofitquarterly.org/policysocial-context/22149-flaws-in-the-social-impact-bond-craze.html

Productivity Commission. (2010). *Contribution of the Not-for-Profit Sector.* Canberra. Retrieved from: http://www.pc.gov.au/projects/study/not-for-profit/report

Productivity Commission. (2004). *Report on Government Services 2004.* Melbourne. Retrieved from: http://www.pc.gov.au/gsp/rogs/2004

Pudelek, J. (2013, June 14), HM Prison Peterborough social impact bond has led to a fall in reconvictions, official figures show. Retrieved from: http://www.thirdsector.co.uk/news/1186265/

Rabin, M. (1993). Incorporating Fairness into Game Theory and Economics. *American Economic Review, 83* (5), 1281-1302.

Rigdon, M. (2002). Efficiency wages in an experimental labor market. *Proceedings of National Academy of Sciences, 99* (20), 13348-13351.

Rushe, D., & Burke, J. (2011, April 18). *Three Cups of Tea author Greg Mortenson in schools fraud row*. Retrieved May 30, 2012, from The Guardian: http://www.guardian.co.uk/books/2011/apr/18/greg-mortenson-three-cups-tea

Schnedler, W., & Vadovic, R. (2011). Legitimacy of Control. *Journal of Economics & Management Strategy, 20* (4), 985-1009

Smith, V. (2002). Method in Experiment: Rhetoric and Reality. *Experimental Economics, 5* (2), 91-110.

Social Finance. (2011). *A Technical Guide to Developing a Social Impact Bond.* Social Finance. London: Social Finance Ltd. Retrieved from: http://www.socialfinance.org.uk/sites/default/files/Technical_Guide_Overview.pdf

Social Finance. (2009). *Social Impact Bonds: Rethinking finance for social outcomes.* London: Social Finance Ltd. Retrieved from: http://payforsuccess.org/sites/default/files/sf_rethinkingfinance.pdf

The Treasury. (2011). A Definition of Charity: Consultation Paper. Commonwealth of Australia 2011. Retrieved from: http://archive.treasury.gov.au/documents/2161/PDF/definition_v6.pdf

Tierney, T. J., & Fleishman, J. L. (2011). *Give Smart. Philanthropy That Gets Results.* New York: Public Affairs Books.

Tirole, J. (1986). Hierarchies and Bureaucracies: On the Role of Collusion in Organizations. *Journal of Law, Economics and Organization, 2* (2), 181-214.

Vignaendra, S., & Fitzgerald, J. (2006, October). Reoffending among young people cautioned by police or who participated in a youth justice conference. *Crime and Justice Bulletin - Contemporary Issues in Crime and Justice* (103). Retrieved from: http://www.lawlink.nsw.gov.au/lawlink/bocsar/ll_bocsar.nsf/vwFiles/CJB103.pdf/$file/CJB103.pdf

Von Glahn, D., & Whistler, C. (2011). Translating plain English: can the Peterborough Social Impact Bond construct apply stateside? *Community Development Investment Review* , 58-70.

Washington State Institute for Public Policy. (2004). *Outcome Evaluation of Washington State's Research-Based Programs for Juvenile Offenders.* Washington State. Retrieved from: http://www.wsipp.wa.gov/rptfiles/04-01-1201.pdf

Wong, J., & Ortmann, A. (2012a). Do donors care about the price of giving? A review of the evidence, with some theory to organize it. Working Paper.

Wong, J., & Ortmann, A. (2012b). A methodology note: On uneven expected earnings in the Lab. Discussion Paper.

Yetman, M., & Yetman, R. (2011). Do Donors Discount Misleading Accounting Information? *UC David Graduate School of Management Research Paper No. 04-09* .

Yetman, M., & Yetman, R. J. (2011a). Do Donors Discount Misleading Accounting Information? *UC Davis Graduate School of Management Research Paper No. 04-09* .

Yetman, R., & Yetman, M. (2011b). The Effects of Governance on the Accuracy of Charitable Expenses Reported by Nonprofit Organizations. *Contemporary Accounting Research*.

Young, D. R. (1981). Entrepreneurship and the Behavior of Nonprofit Organizations: Elements of a Theory. In M. White (Ed.), *Nonprofit Firms in a Three-Sector Economy* (pp. 135-162). Washington DC: The Urban Institute Press.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, 13(1), 75-98.