Optimal incentives on multiple prosocial activities when reputation matters*

Francisco Candel-Sánchez[†] Juan Perote-Peña[‡]

Abstract

In this paper we present a signaling model in which individuals engage in socially beneficial but costly activities to convey information about their willingness to cooperate with other agents. When several activities are available, the inclusion of monetary compensations in anyone of them affects the relative costs of undertaking each activity and, therefore, their informative value for agents. We find the subsidies that maximize social welfare, which are shown to depend critically on the reputation gained from each activity. Finally, we use comparative statics analysis to study the effects on optimal subsidies of changes in their determinants.

*We thank Jorge Navarro for helpful advice on doubly truncated distributions. We also thank three anonymous referees and Stephen Hasler for helpful comments and suggestions. Financial support from the Spanish Ministry of Economy and Competitiveness through MEC/FEDER grant ECO2016-75631-P is gratefully acknowledged.

[†]Universidad de Murcia. Murcia, 30100, Spain. E-mail: fcandel@um.es.

[‡]Universidad de Zaragoza. Zaragoza, 50009, Spain. E-mail: jperote@unizar.es.

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/sjoe.12369

Keywords: Signaling; economic incentives; social welfare.

JEL codes: D03; D78; D82; H31; H41.

1 Introduction

Prosocial activities, such as volunteering, blood donations or giving to charities, are essentially privately provided public goods, and donors face a Prisoner's Dilemma situation. Public decision makers often establish subsidies to encourage the provision levels of activities that generate positive externalities, but the empirical evidence suggests that material incentives can backfire in practice, inducing partial or total crowding out of prosocial behavior[1]. Several psychological mechanisms have been proposed in the literature to account for this; Bowles and Polanía-Reyes (2012) argue that "incentives may (i) provide information about the person who implemented the incentive, (ii) frame the decision situation so as to suggest appropriate behavior, (iii) compromise a control averse individual's sense of autonomy, and (iv) affect the process by which people learn new preferences". In this paper, we obtain a particular crowding-out effect for a new reason: since agents may use a number of mutually exclusive prosocial activities to signal their degree of altruism credibly, monetary incentives will change their relative costs (or prices) and, therefore, cause shifts in behavior that may lead to a lower aggregate level of all prosocial activities.

Imagine that a local government is considering subsidizing two volunteer activities: (1) conservation (repairing paths, clearing ponds and waterways or planting trees), and (2) caregiving (providing assistance and support to people with developmental needs, eg. helping

[1] See Frey and Jegen (2001) for a survey, and Gneezy, Meier and Rey-Biel (2011) for an analysis of the contexts where incentives (don't) work to modify behavior.

people with learning disabilities). Both types of prosocial activities are publicly visible, and undertaking either of them improves the image and reputation of the volunteers. Now, imagine that the local government seeks to promote forest clearing by offering a subsidy. Although the new incentive encourages people to participate in conservation activities, it might also reduce the amount of volunteer caregiving, even to the extent of decreasing the total amount of volunteering. This paradoxical result can be attributed to cross effects arising when subsidies for (potentially competing) prosocial activities change their relative benefits.

There are many real-world examples where material incentives for a prosocial activity crowd out other similar and related prosocial activities. For instance, Robért and Jonsson (2006) find that a free public transport policy does not substantially reduce private vehicle use. Instead, this policy crowds out other prosocial modes of transport such as walking or cycling. Similarly, economic incentives on electric vehicles, although replacing conventional vehicles to some extent, also increase overall car transport and substitute public transport and other means of transport[2]. Kits et al (2014) show that the introduction of incentives for conservation activities (in the form of conservation auctions) reduces monetary donations to an environmental charity. Lilley and Slonim (2014) find that time donations (volunteering) and monetary donations are likely to be net substitutes, and that a matching donation is effective in increasing monetary donations, but also causes a substitution away from volunteering. In the context of solid waste management and recycling, empirical studies by Kinnaman and Fullerton (2000) and Hong (1999), among others, show that collection fees

[2]See, for instance, Franke et al. (2012), Figembaum et al. (2014), Halvorsen and Frøyen (2009), or Holtsmark and Skonhoft (2014).

on garbage disposal may increase illegal dumping or burning.

In this paper we propose a stylized model to account for these effects. We first characterize the optimal structure of incentives on prosocial activities and then explore its sensitivity to certain changes in the system. In order to provide a rationale to cross crowding out effects we consider that the agents are heterogeneous in their degree of social preferences. Since the agent's willingness to cooperate (the "type") is private information, individuals use prosocial activities as an instrument to signal their types credibly. Even though agents do not have preferences about prosocial activities per se, they are concerned with the reputation associated with each activity. Hence, in the separating equilibrium of a signaling game, activities (with different costs) are classified according to the reputation benefits they entail for individuals. We show that incorporating explicit rewards in one activity changes the relative costs of all prosocial activities, and this induces substitution effects among activities. This is because the change in relative costs modifies the information conveyed through the choice of each action. Prosocial activities can therefore be interpreted as "competing signals" in the sense that they are alternative channels for signaling credible information about reputation.

The explicit consideration of multiple activities produces a number of insights that could not have been obtained otherwise. Consider the example in the second paragraph of this section, and assume (w.l.o.g.) that conservation (activity 1) is more costly to the agent than caregiving (activity 2). In equilibrium, conservation activities provide a high reputation and caregiving activities provide a moderate reputation. A subsidy on conservation activities displaces the highest-type caregivers to conservation. This elimination of the best among the caregivers reduces the reputational benefit to caregiving. This leads to the lowest-type caregivers to abandon caregiving. The set of people who don't contribute thus grows.

Reputation spillovers from the targeted activity to other related activities are the key to explain cross crowding out effects. Then, if the government's goal is to increase the total amount of both activities, either the subsidy on conservation activities must be lowered or the subsidy for caregiving increased, or both.

The comparative statics analysis on optimal subsidies also offers a number of seemingly counter-intuitive results. For example, if the cost of conservation activities increases, it is optimal to lower subsidies for both activities. Moreover, if conservation activities become more popular or socially valued, it is optimal to increase the subsidy for voluntary caregiving too. Rather surprisingly, a higher valuation of caregiving activities may induce lower subsidies for them. It should be noted that the cross crowding-out effects between activities identified in this paper are far from being specific or pathological cases. On the contrary, they take place under reasonably general conditions. Thus, our results point to the importance of considering these new complexities in the design of optimal incentives by policy-makers.

The paper is organized as follows. Section 2 offers a brief background of the related literature. Section 3 sets up the basic framework of analysis and presents the individuals' utility functions. Section 4 studies the equilibrium choice of the agents for a given subsidy scheme. Section 5 analyzes the impact of subsidies on the supply of prosocial activities, which allows us to identify the crowding-out effects across activities. In Section 6, after an appropriate definition of social welfare, we characterize the optimal subsidy profile. In Section 7, we perform comparative statics analysis to ascertain how optimal subsidies change in response to exogenous changes in their determinants. Section 8 presents the main conclusions of the paper. The proofs of the main results in the text can be found in Appendix A (online). Appendices B and C (also online) are devoted to analyzing examples, extensions and robustness of our model.

2 Related literature

In this paper we study a problem in which a social planner must devise a subsidy scheme for multiple prosocial activities used by the agents as alternative signaling devices. This problem is related to the well-known multitask problem, pioneered in a seminal paper by Holmstrom and Milgrom (1991) and used by other authors since then[3]. The multitask approach is developed in a principal-agent setting, mainly in the context of firm organization. A first difference of the multitask approach, with respect to our model, is that the agent's actions (typically effort) are not perfectly observable, leading to a moral hazard problem, whereas in our approach the performance of prosocial activities is public information. In fact, the agents use these activities to signal their willingness to cooperate. Secondly, in a multitask approach, the agent's cost depends on the total effort devoted to all tasks, which can then be complements or substitutes. The effort exerted by the agent in one task may crowd out the other task because the tasks are not technologically independent. Instead, in our model, agents can only choose one action from a finite set of alternatives (e.g., recycling garbage vs. waste deposits in bins and bags, or public transport vs. bicycle). As in the multitasking approach, our model includes standard substitution effects due to changes in relative prices. This is why, for instance, a subsidy on electric vehicles (EVs) displaces some purchases from conventional vehicles to EVs. However, unlike the multitasking approach, this change in relative prices also affects the reputational value of all modes of transport. Thus, while some people switch from conventional cars to electric cars (this is the desired effect of the policy),

[3] For a survey on multitask agency theory, see Dewatripont et al. (2000).

electric vehicles also substitute other environmentally friendly modes of transport, such as public transport, walking or cycling. As will become apparent in Section 5 of the paper, crowding out occurs because subsidies affect the agents' reputational payoffs by changing the equilibrium partition of their types.

In the context of prosocial activities, our multi-activity approach is related to Ek (2017, 2018), who estimates the magnitude of cross-price effects across prosocial alternatives. This author finds evidence that substitution decreases as activities become more dissimilar. Specifically, in Ek (2018), it is shown that there exist sufficient conditions under which facilitating one activity crowds out effort in other activities. One major difference from our model is that Ek (2018) does not account for the agents' concern with reputation. A signaling model with multi-dimensional actions is also used in Ellingsen and Johannesson (2011) to study the choice of monetary vs. non-monetary gifts. Individuals are either selfish or altruistic and they receive indirect utility from social esteem. The main difference with our model is that in Ellingsen and Johannesson (2011) agents receive some (compassion) utility from the opponent's consumption level.

One branch of social psychology literature studies the effect of explicit incentives on individuals' "intrinsic motivation" to undertake certain activities. Extrinsic motivation (e.g. monetary incentives) may displace intrinsic motivation, even to the extent of being counterproductive. Some authors have dealt with this by assuming a reduced-form crowding-out function of intrinsic motivation (e.g., Frey and Oberholzer-Gee, 1996). Recent approaches to understanding the Motivation Crowding-Out Effect involve endogeneizing, rather than assuming, the way individuals behave when faced with explicit rewards or penalties. Contributions by Bénabou and Tirole (2006) (B & T from now on) and Seabright (2009) emphasize the informational aspects of crowding-out effects using models where rational agents decide on their participation in prosocial activities according to reputation concerns.

In B & T each agent is endowed with private information, expressed as the agent's bi-dimensional type, with one dimension accounting for altruism and the other for greed. Rewards on a prosocial activity introduce noise to the signal-extraction problem of the agent's type. Under a certain range of parameters, and assuming that types are normally distributed, B & T identify instances in which incentives for the social activity reduce its supply. Very related to B & T, Seabright (2009) analyzes a screening context in which agents decide to participate in a civic activity or not after some reward for participating has been announced. This author considers a sorting condition that establishes that the agents' expected gain from a profitable assortative matching in the future is increasing in the agent's type.

Unlike B & T, we consider one-dimensional types, and the social activities can be subsidized separately. The (rather restricting) sorting condition in Seabright (2009) is not required in our model, but we must consider at least two prosocial activities to obtain our cross crowding out effects. Both in B & T and Seabright (2009) there is a trade-off between the direct effect of rewards and their indirect effects on the agent's reputation and/or intrinsic motivation. The agents' equilibrium choice in the face of this trade-off involves, under some circumstances, a lower supply of the social activity that has been incentivized. In contrast, the mechanism that drives crowding out in our paper is based on substitution effects among different activities, whose relative cost can be modified by subsidies. In our framework, rewards for a given activity affect the information that agents reveal when undertaking other activities. This makes it possible, for instance, to affect the aggregate supply of activity A by subsidizing activity B. This property is particularly relevant in decision-making when a subsidy on A is unfeasible due to legal restrictions or high transaction costs.

Consider the empirical evidence regarding the Norwegian policy of subsidizing EVs. In B&T's model, subsidies on EVs spoil the signaling value of acquiring an EV to the extent that the purchases of EVs may even decrease. In our model, instead, the lower cost of EV's after the subsidy induces people to replace conventional cars by electric cars (the price effect predominates over the reputation effect on the subsidized activity). However, the change in relative prices also affects the reputation associated to other modes of transport. The empirical evidence reveals a displacement of other good activities like walking, cycling or public transport, jointly with an increase in overall car use. A major difference with B&T, then, is that our model includes reputational spillovers among activities. Such effects become apparent also in other contexts. For instance, as shown in Kinnaman and Fullerton (2000), the introduction of collection fees on waste disposal aimed to encourage recycling activities has the undesirable side effect of increasing illegal dumping or burning.

The standard public economics approach to the analysis of subsidies (taxes) relies on the general principle that agents respond to incentives. Governments use subsidies to promote activities that generate positive externalities, or taxes to discourage activities that cause negative externalities. Bowles and Hwang (2008) combine the public economics approach with the analysis of individual behavior when faced with explicit incentives, like the Motivation Crowding Out effect. These authors investigate the design of optimal incentives in a context where economic incentives and social preferences can be either complements or substitutes and the separability assumption[4] does not hold. A notable difference with the present paper is that they use a given "black box" function to model the trade-off between extrinsic and intrinsic motivation. Our approach is based on a signaling game in which individuals' responses to incentives are endogenous, and intrinsic motivation is not considered. Instead, citizens understand how explicit incentives modify the information revealed through the performance of each activity and make rational and self-interested decisions based on that. The reason why explicit incentives crowd-out prosocial behavior in our model is not because they provoke a displacement of intrinsic motivation, but because they change the reputation benefits associated to each activity.

Finally, this paper is also related to Greenwald and Stiglitz (1986), who study Pareto-improving taxes/subsidies in economies with incomplete markets and imperfect information. In analyzing the welfare properties of signaling equilibria, these authors find that the government can establish taxes or subsidies on goods that change the extent of signaling and can lead to welfare improving allocations.

3 Model

We develop a signaling model in which a continuous set of heterogeneous agents (or individuals) undertake prosocial activities observable by all of them. Agents use this observed behavior to estimate the other agents' types, and the final payoff of each agent is highly affected by the estimates about their own type made by other agents.

Let us denote the agent's type by $\theta_i \in [0, 1]$. It represents a somehow measurable personal

[4] The separability assumption is implicitly used in the public choice literature and establishes that the level of material rewards or penalties does not affect the agents' public goods valuations.

characteristic or trait of agent i, and is unobservable directly by agents other than i, that is, each agent's type is her only private information in the game. Agents' types are distributed in the interval $[0,1]$ according to cumulative distribution function (cdf) $F(\theta)$. Therefore, given any $x \in [0,1]$, $F(x)$ represents the fraction of the population with types lower than or equal to x. Each agent is allowed to undertake one (and only one) action from the discrete set $A = \{a_0, a_1, a_2\}$. We denote as c_j the objective cost of carrying out action a_j, for $j = 0, 1, 2$, and we assume $c_1 > c_2 > c_0 = 0$. Actions a_1 and a_2 are interpreted as costly prosocial activities. In turn, choosing the costless action, a_0, means that the agent does not undertake any prosocial activity. We also assume that individuals cannot decide about their degree of participation in the activities chosen: they can only choose one of the three options available to them[5].

Types can be interpreted as the agents' degrees of cooperative behavior or willingness to contribute to a public good (e.g. recycling household waste, participating in voluntary social work, etc.), or as the agent's degree of altruism. In general, we talk about the individual's *willingness to cooperate*. Individuals are concerned with the perceptions that others have about their own types, so showing high reputation leads to a higher payoff by obtaining better cooperative matches in future social interactions[6].

Since some agents can enhance their perceived reputation by performing costly activities, activities 1 and 2 may be used as revealing signals. We assume that the cost to agent i of

[5] This setting is adequate to explain the evidence of *dugnads* in Brekke et al (2003) or the evidence recorded in day care centers in Gneezy and Rustichini (2000).

[6] In a different context, Rege (2008) considers a *complementary interaction* process that induces the agents to care about social status, since the investment in status (to buy a Rolex watch) serves as a signal of non-observable abilities (high business skills).

sending signal a_j is negatively correlated with agent $i's$ type, so the subjective cost from undertaking any given prosocial activity is lower for those individuals who are themselves more cooperative. For the sake of simplicity, we propose the following linear form for the cost[7] of activity j to agent i : $C(a_j \mid \theta_i) = (1-\theta_i)c_j$. For instance, the cost of a_1 for an individual of type $\theta_i = 0.7$ is equal to $0.3c_1$, while an individual whose type is $\theta_i = 0.1$ faces a much higher cost of $0.9c_1$.

Prosocial activities are considered pure public goods as long as their benefits to society are non-rival and non-excludable. Individuals also benefit from other types of public goods that cannot be provided individually. Let us denote by A_j the aggregate supply of prosocial activity $j = 1, 2$, and let g stand for the expenditure on other public goods. Agent $i's$ benefit is denoted as $V_i(A_1, A_2, g)$, where a fraction, A_j, of agents take prosocial action $j = 1, 2$, and other goods are publicly provided at level g. Note that, given the infinite population considered, all individuals have zero weight in the aggregate levels of activities[8]. Therefore, from the point of view of agent i, her benefit $V_i(A_1, A_2, g)$ is exogenous, since it cannot be significantly affected by her individual choices. We denote it simply by V_i.

Let us consider the introduction of a subsidy scheme $s = (s_1, s_2)$, where $s_j \geq 0$ stands for

[7] In fact, all our results can be generalized to any differentiable cost function such that the difference $c(a_1, \theta_i) - c(a_2, \theta_i)$ is strictly increasing in the type θ_i. Since it is relatively cheaper for cooperative individuals to produce the costly signal a_1, the difference $u_i(a_1/\theta_i, s) - u_i(a_2/\theta_i, s)$ is strictly increasing in θ_i. This property, known as Increasing Differences (Topkis, 1978), is a version of Spence-Mirrlees well known "single-crossing" condition, which plays a prominent role in signaling games.

[8] In this sense, prosocial behavior closely resembles the action of voting in Anthony Downs' "voting paradox": Each voter faces a positive cost of casting a vote, but the voter's influence on the electoral outcome is close to zero.

the subsidy[9] established for activity $j = 1, 2$. The policy-maker, who faces budget constraint $B \geq s_1 A_1 + s_2 A_2 + g$ (with B being exogenous), may want to establish a subsidy scheme, s, in order to encourage the agents to participate in prosocial activities. Given s, each agent in $[0, 1]$ chooses one (and only one) activity in set A. We consider a signaling game in which each agent plays the role of "the sender" when choosing a_j and the rest of the population is receiver of the signal that updates their beliefs on the sender's type after the signal has been observed. The timing of this game is as follows: (i) each agent selects an element of set A; (ii) all agents observe the other agents' choices, and (iii) the payoffs are realized. We define the strategy of agent i under subsidy scheme s as a function $\sigma_i(\theta_i, s) : [0, 1] \times R^2 \rightarrow A$. The atomistic representation of the population implies that no single agent can affect other agents' payoffs, and therefore strategic interactions are absent from the analysis. However, each individual's choice is publicly observed and affects her own reputation.

We define the reputation function of agent i, who undertakes activity a_j under subsidy scheme s, as $R_i(a_j, s) = E[\theta_i \mid a_j, s]$, i.e., the mathematical expectation of the agent's type conditional on a_j, given s. This expectation accounts for the private returns of building a "solid reputation" provided that all agents other than i (receivers) update their beliefs on agent $i's$ type (the sender) after observing activity a_j (the signal). These private benefits can be interpreted as the discounted payoffs from profitable assortative future matching with

[9] In general, in this paper we analyze the effect of incentives on multiple prosocial activities. Such incentives can be positive (subsidies, if $s_j > 0$), or negative (taxes, if $s_j < 0$). For easier interpretation of the results, though, we speak about subsidies throughout the paper.

other cooperative agents[10].

Agent $i's$ preferences can be expressed by the following additively separable utility function:

$$u_i(a_j \mid \theta_i, s) = R_i(a_j, s) - C(a_j \mid \theta_i) + I_i + V_i, \quad (1)$$

where the term I_i represents income received by agent i (in this model, the monetary amount of subsidies received). The linear form of function u_i means that we are implicitly assuming that the agent's valuation of money is independent of her attitude towards prosocial behavior[11].

We are interested in studying the effects that the introduction of subsidies have on the aggregate supply of activities 1 and 2. For this purpose, we first need to analyze the influence of subsidies on the agents' equilibrium choice.

4 Individuals' equilibrium choices

The subsidy profile $s = (s_1, s_2)$ changes the relative cost of activities, thus modifying the informative value of each signal with regard to individuals' willingness to cooperate. In this section we analyze the influence of those subsidies on the agents' equilibrium choices. After a monetary subsidy $s_j \geq 0$ is introduced on activity $j = 1, 2$, the (subjective) cost of activity a_j to agent i reduces to $C(a_j \mid \theta_i) - s_j$.

Individuals are considered rational and their preferences are given by Eq. (1). Posterior

[10] Seabright (2009) explicitly considers a process in which efficient matchings are characterized after individuals choose to participate or not in a civic activity. Our reputation function is very similar to the agent's reward considered in Dewatripont et al (1999) in a multitask problem.

[11] We introduce altruism linearly in the same fashion as Buurman and Dur (2012).

beliefs on agent $i's$ type are formed after action a_j is observed. Let $f(\theta_i)$ be the prior density function over the types. We define as $\mu_{ij}(s)$ the belief regarding agent $i's$ type after everybody observes that agent i undertakes activity a_j, given the current subsidy scheme s. Namely, $\mu_{ij}(s) = f(\theta_i \mid a_j, s)$.

The agents' payoffs and strategies, the set of possible types and the prior distribution function of the types define a dynamic Bayesian game. In order to solve this game, we use the Perfect Bayesian Equilibrium concept, and refer to it simply as "equilibrium". In equilibrium, beliefs are updated according to Bayes' rule given the agents' actions, and each agent chooses the action that maximizes her utility given these beliefs.

We focus our attention on a semi-separating equilibrium in which those agents with higher types undertake activity 1, intermediate types undertake activity 2 and the lower types choose not to carry out any activity. Let us then consider a partition of interval $[0, 1]$, with $\Theta_j \subset [0, 1]$ being the set of agents who choose action a_j, for $j = 0, 1, 2$. Note that the equilibrium characterized here is not fully separating, i.e., the choice of an action does not reveal the exact true value of θ_i. Instead, the set of agents is partitioned into three subsets, and each agent is pooled with all the other agents who belong to the same subset. Information is not fully revealed because the agents are only allowed to choose from a discrete set of actions[12].

Definition 1: *An equilibrium is a set of strategies* $\sigma_i^*(\theta_i, s)$ *and beliefs* $\mu_{ij}^*(s)$*, for all* i

[12] Technically, we compute a semi-pooling equilibrium, in which some types of agents choose the same action, while other types choose different actions. A pure separating equilibrium is not possible when there is a continuum of types but a discrete set of signaling actions.

and for $j = 0, 1, 2$ *such that*

$$\sigma_i^*(\theta_i, s) = \arg\max_{\{a_j\}} u_i(a_j \mid \theta_i, s) = \begin{cases} a_1 & \text{if} \quad \theta_i \in \Theta_1 \\ a_2 & \text{if} \quad \theta_i \in \Theta_2 \\ a_0 & \text{if} \quad \theta_i \in \Theta_0 \end{cases}$$

and

$$\mu_{ij}^*(s) = \begin{cases} f\left(\theta_i \mid \sigma_i^*(\theta_i, s), s\right) & \text{if} \quad \theta_i \in \Theta_j \\ 0 & \text{if} \quad \theta_i \notin \Theta_j \end{cases}$$

According to the beliefs in Definition 1, an individual who undertakes a_1 must belong to set Θ_1. A similar inference is made for individuals who choose a_2 and a_0. We further adopt the Intuitive Criterion (Cho and Kreps, 1987) as a refinement on beliefs in order to rule out pooling equilibria[13].

The reputation function of agent i with type $\theta_i \in [0, 1]$, evaluated at the equilibrium strategy $\sigma_i^*(\theta_i, s)$, is given by

$$R_i(\sigma_i^*(\theta_i, s), s) = E\left[\theta_i \mid \sigma_i^*(\theta_i, s), s\right] = \frac{\displaystyle\int_{\theta_i \in \Theta_j} \theta_i f\left(\theta_i\right) d\theta_i}{\displaystyle\int_{\theta_i \in \Theta_j} f\left(\theta_i\right) d\theta_i}. \tag{2}$$

Lemma 1: Given the equilibrium reputation associated to each action choice, the partition of interval $[0, 1]$ into sets Θ_1, Θ_2 and Θ_0 is characterized as $\Theta_1 = [\alpha, 1]$, $\Theta_2 = [\beta, \alpha]$

[13]Consider a trivial equilibrium in which each agent chooses a_0 regardless of his type and $\mu_{ij}(s) = f(\theta_i)$ for each a_j, i.e., the agents' actions do not affect their reputations. Given these beliefs, choosing action a_0 is optimal since this action is costless and the other (costly) actions have no influence in $V_i(A_1, A_2, g)$. This equilibrium would not satisfy the Intuitive Criterion: an agent with low type is never interested in deviating to action a_2 or a_1. However, such deviations may be profitable for higher types for beliefs other than those assumed in the trivial pooling equilibrium. Then, if some agent deviates from choosing a_0, she must be of high type, and it is "not intuitive" that this agent does not improve her reputation by choosing a_2 or a_1.

and $\Theta_0 = [0, \beta]$, where α is the unique type of agent indifferent to choosing a_1 or a_2, and β is the unique type indifferent between a_2 and a_0.

Proof: See Appendix A.

The information inferred from each action is summarized in the average type corresponding to each interval (i.e., reputation). In an interior equilibrium[14], thresholds α and β must be such that $0 \leq \beta \leq \alpha \leq 1$. Observe that thresholds α and β are parameterized by the subsidy scheme s. For notational simplicity, though, we shall omit s as an argument of α and β throughout the paper. We write the equilibrium reputation in Eq. (2) of an agent who undertakes activity a_1 as $R_i(a_1, s) = r_1(\alpha)$. Similarly, the terms $r_2(\alpha, \beta)$ and $r_0(\beta)$ refer to the reputation of individuals who take actions a_2 and a_0, respectively. Note that, under our conditions, the reputation functions $r_1(\alpha)$, $r_2(\alpha, \beta)$ and $r_0(\beta)$ are continuous and strictly increasing in their arguments (see Balakrishnan, 2001). For the sake of simplicity, in the analysis that follows we assume that they are also differentiable. We denote as $r_{j\alpha}$ and $r_{j\beta}$ (with $j = 0, 1, 2$) the partial derivatives of functions $r_j(.)$ with respect to α and β, respectively. Then, it holds that $r_{1\alpha} > 0$, $r_{2\alpha} > 0$, $r_{2\beta} > 0$ and $r_{0\beta} > 0$. It is useful to define $\Delta r_\beta = r_{2\beta} - r_{0\beta}$ and $\Delta r_\alpha = r_{1\alpha} - r_{2\alpha}$ as the relative reputation gains from variations in thresholds β and α, respectively.

Lemma 2: If the agents' types are distributed according to a concave cdf $F(.)$, then it holds that $\Delta r_\alpha \geq 0$ and $\Delta r_\beta \geq 0$.

Proof: See Appendix A.

Conditions $\Delta r_\beta \geq 0$, $\Delta r_\alpha \geq 0$ have a fairly simple interpretation: $\Delta r_\beta \geq 0$ means that,

[14] In Appendix B (online) we discuss the existence of other possible semi-separating equilibria which may arise under certain parameter configurations.

if some individuals (those at the bottom of set Θ_2) abandon activity 2 and select a_0 instead (i.e., β increases), the reputation gain of individuals who remain in set Θ_2 is greater than or equal to the reputation gain for individuals in set Θ_0. A similar interpretation can be given to condition $\Delta r_\alpha \geq 0$. In the rest of the paper we assume that $F(.)$ is concave. In order to obtain an interior equilibrium in the next proposition, we also assume that the cost c_2 is high enough relative to subsidy s_2.

Proposition 1: *If the cdf of the types, $F(.)$, is concave and the cost c_2 is high enough relative to subsidy s_2, an interior equilibrium (α, β) exists where α and β are fully determined by the following two equations:*

$$r_1(\alpha) - (1-\alpha)c_1 + s_1 = r_2(\alpha,\beta) - (1-\alpha)c_2 + s_2 \tag{3}$$

$$r_2(\alpha,\beta) - (1-\beta)c_2 + s_2 = r_0(\beta) \tag{4}$$

Proof: See Appendix A.

For any given pair (c_1, c_2), Eqs. (3) and (4) in Proposition 1 implicitly define thresholds α and β as a function of subsidies s_1 and s_2.

5 Cross crowding-out effects

Once we have derived the equilibrium choices of the agents for a given subsidy scheme s, we are interested in the local response of thresholds α and β to small changes in the subsidy on activity 1 ($ds_1 > 0$) and/or in the subsidy on activity 2 ($ds_2 > 0$). In a separating equilibrium, like the one characterized above, the level of activity 1 is $A_1(\alpha) = \int_\alpha^1 f(\theta_i)d\theta_i$, the level of activity 2 is $A_2(\alpha,\beta) = \int_\beta^\alpha f(\theta_i)d\theta_i$ and the aggregate level of prosocial activities[15] is given by $A_1(\alpha)+A_2(\alpha,\beta) = \int_\beta^1 f(\theta_i)d\theta_i$. Observe that the signs of the following derivatives hold for any density function $f(\theta_i)$: $A_1' < 0$, $\frac{\partial A_2}{\partial \alpha} > 0$, $\frac{\partial A_2}{\partial \beta} < 0$, $\frac{\partial (A_1+A_2)}{\partial \beta} < 0$ and $\frac{\partial (A_1+A_2)}{\partial \alpha} = 0$.

In this section we characterize cross crowding-out effects as follows: (i) subsidizing the most costly activity reduces the supply of the least costly activity and the total supply of prosocial activities; and, (ii) raising the subsidy on the least costly activity may affect the supply of the most costly activity negatively.

Next we analyze the influence of subsidies on the supply of prosocial activities. We denote by α_j (β_j) the derivative of threshold α (β) with respect to subsidy s_j, with $j = 1,2$. We resort to implicit differentiation in Eqs. (3) and (4) to obtain

$$\beta_1 = \frac{r_{2\alpha}}{D}, \tag{5}$$

$$\alpha_1 = -\frac{\Delta r_\beta + c_2}{D}, \tag{6}$$

$$\beta_2 = -\frac{r_{1\alpha} + c_1 - c_2}{D}, \tag{7}$$

$$\alpha_2 = \frac{c_2 - r_{0\beta}}{D}, \tag{8}$$

with $D = r_{2\alpha}r_{2\beta} + (\Delta r_\beta + c_2)(\Delta r_\alpha + c_1 - c_2)$.

Note that Lemma 2 implies that $D > 0$, so we can obtain a clear sign of the derivatives above.

Proposition 2: *(i) A subsidy on activity 1 increases the supply of activity 1* $(\alpha_1 < 0)$*, but also reduces the aggregate supply of prosocial activities* $(\beta_1 > 0)$ *and the supply of activity*

[15] By construction, in our model we have that $A_1 + A_2 + A_0 = 1$. An alternative interpretation of the model would be to consider that there is a single agent with possible types drawn from the interval $[0,1]$, and then A_j would be the probability that this individual chooses action j given profile s.

2 $(\alpha_1 - \beta_1 < 0)$; *(ii) A subsidy on activity 2 increases the total supply of prosocial activities* $(\beta_2 < 0)$, *but it crowds out activity 1* $(\alpha_2 > 0)$ *if* $c_2 > r_{0\beta}$.

Proof: Straightforward from Eqs. (5)-(8).

In order to develop some intuition about crossed crowding out effects, suppose that s_1 increases. In this case, the lower cost of activity 1 after the subsidy induces the most cooperative agents in set Θ_2 to shift to activity 1, thus worsening the reputation of the agents who keep doing activity 2. This reputation loss causes some individuals (those at the bottom of set Θ_2) to abandon activity 2 and select a_0 instead. The equilibrium supply of activity 2 decreases for two reasons: some people substitute it by activity 1 (the highest types in Θ_2) and other people substitute it by activity 0 (the lowest types in Θ_2). On the other hand, if s_2 increases, the condition that ensures crowding out effects on the supply of activity 1 $(\alpha_2 > 0)$ is $c_2 > r_{0\beta}$. The effect of an increase in s_2 on the supply of activity 1 is, in general, ambiguous because it is the combination of two opposite effects: a direct price effect whereby activity 1 becomes more expensive with respect to the subsidized activity 2 (this would, *ceteris paribus*, reduce the supply of activity 1) and a reputation effect that makes activity 2 less attractive for the types at the top of set Θ_2.

Our results in Proposition 2 establish that, in general, incentives may backfire due to cross crowding out effects. If the government uses a subsidy to boost the level of activity 1, it may cause a reduction of the level of activity 2 to the extent of leading to an overall reduction of all prosocial activities. Besides, a subsidy on activity 2 may crowd-out the supply of activity 1 if the cost of activity 2 is high enough[16]. The singularity of our approach lies in the fact that a change in relative prices displaces agents between activities and this

[16] Specifically, if the distribution function is concave, for $c_2 > r_{0\beta}$ to hold, it is sufficient that $c_2 \geq 1$.

displacement affects the reputation associated to each activity. The equilibrium partition of individuals, determined by thresholds α and β, is ultimately affected by subsidies through a combination of (direct) price effects and (induced) reputation effects.

There is consistent evidence that subsidies on one prosocial activity sometimes serve to reduce other prosocial activities. For instance, in the context of household solid waste management and recycling, Kinnaman and Fullerton (2000) estimate that a \$1 user fee could decrease the quantity of garbage by 412 pounds per person per year but increase recycling by only 30 pounds per person per year. These authors suggest that the extra garbage might lead to illegal dumping. Similarly, Hong (1999) concludes that a unit pricing system provides pervasive incentives for households to dump or incinerate wastes illegally. Consider three possible actions: Recycling waste (activity 1), traditional household waste in bins and bags (activity 2) and illegal dumping or burning (activity 0). In the terms of our model, a waste fee is equivalent to a reduction of the subsidy on activity 2. According to Proposition 2, it must induce a lower level of prosocial activities (an increase in β) and also a higher rate of recycling (a decrease in α) when $c_2 > r_{0\beta}$. The empirical evidence bears out this theoretical prediction.

Purchasing an electric vehicle (EV) is a signal of environmentally friendly behavior, and thus it influences the agent's reputation positively (Ariely et. al (2009)). Consider now the following three modes of transport: Public transport (activity 1), use of an EV (activity 2) and use of a conventional vehicle (activity 0). Over the last few years, Norway has implemented a generous policy of subsidies and exemptions aimed to encourage the purchase and use of EV's. However, as stated in Holtsmark and Skonhoft (2014), this policy may not have achieved the desired goals. A report by Halvorsen and Froyen (2009) concludes that people reduce their use of public transport for commuting to work after purchasing an EV. Before the acquisition of the EV, they take 23% of their trips to work by public transport. This percentage decreases to 6% after they buy the EV. Using data from a field study in the Berlin metropolitan area, Franke et al (2014) report that 29.5% of EV users choose public transport, bike or foot as modes of transport before they receive an EV, while this percentage falls to 8.9% after receiving an EV. Figembaum et al (2014) found that, after acquiring an EV, the use of public transportation fell by 24% of the cases, while only 4% increased their use.

The results in Proposition 2 suggest that if the policy-maker is interested in stimulating some activity through subsidies, she must take into account the presence of cross crowding-out effects which may induce undesired (or at least unexpected) consequences on the supply of other activities. The government's behavior should also be shaped by its own valuation of prosocial activities compared to the opportunity costs of subsidies. This raises a relevant public policy issue: If the government is aware that cross crowding out effects occur when agents are concerned with reputation, how should an optimal subsidy scheme be designed? Moreover, how would such a scheme respond to changes in its determinants? In the next section we address both questions.

6 Optimal subsidies on prosocial activities

We use the model presented in the previous sections to characterize a subsidy profile $s^* = (s_1^*, s_2^*)$ that maximizes social welfare. Then we analyze the responsiveness of s^* to changes in the parameters of the model. Our analysis is carried out under the assumption that the agents adopt the (separating) equilibrium strategies introduced in the previous sections and that the policy-maker or government can use subsidies as an instrument to increase social welfare.

In order to define social welfare properly, we must take into account the budgetary costs of subsidies for the government. Let us recall that a given public budget, B, is devoted either to providing an amount g of public goods, or to subsidizing activities 1 and/or 2. Therefore, the government faces the following budget constraint[17]:

$$g + s_1 A_1 + s_2 A_2 \leq B. \tag{9}$$

We adopt the standard utilitarian approach and define social welfare as the sum of individuals' utilities at equilibrium, $W(s,g) = \int_0^1 u_i(\sigma_i^*(\theta_i, s) \mid \theta_i, s) d\theta_i$. Social welfare consists of four elements: (i) Aggregate reputation; (ii) aggregate costs from prosocial activities; (iii) aggregate benefits from prosocial activities and other public goods, and (iv) aggregate income:

$$\begin{aligned} W(s,g) &= \int_0^1 R_i(\sigma_i^*(\theta_i, s), s) d\theta_i - \int_0^1 C(\sigma_i^*(\theta_i, s) \mid \theta_i) d\theta_i + \\ &+ \int_0^1 V_i(A_1, A_2, g) d\theta_i + \int_0^1 I_i d\theta_i. \end{aligned} \tag{10}$$

Note that aggregate reputation does not depend on s for any distribution function of the types. In fact, aggregate reputation is equal to the average type of an individual in $[0,1]$, which we call R. To see this, notice that

$$\int_0^1 R_i(\sigma_i^*(\theta_i, s), s) d\theta_i = \sum_{j=0}^{j=2} \int_{\theta_i \in \Theta_j} \theta_i f(\theta_i) d\theta_i = \int_0^1 \theta_i f(\theta_i) d\theta_i = R. \tag{11}$$

By construction of the model, wherever the thresholds α and β are located on interval $[0,1]$, and for any density function of the types, the sum of the expected types of the agents who

[17] We assume that budget B comes from non-distortionary taxation on the citizens, and we do not consider the possibility that the government incurs in budget deficit.

belong to sets Θ_0, Θ_1 and Θ_2 is always equal to R. Reputation is zero (constant) sum because, as in a positional game, reputation gains and losses are always compensated in aggregate terms. However, the fact that agents compete for reputation plays a fundamental role in their motivation for prosocial behavior. Actually, in this model, reputation building is the only driving force of prosocial activities.

The aggregate cost of prosocial activities is given by

$$\int_0^1 C(\sigma_i^*(\theta_i, s) \mid \theta_i)d\theta_i = c_1 \int_\alpha^1 (1-\theta_i)f(\theta_i)d\theta_i + c_2 \int_\beta^\alpha (1-\theta_i)f(\theta_i)d\theta_i = C(\alpha, \beta, c_1, c_2). \quad (12)$$

Note that the cost function above depends on both the separating equilibrium thresholds α and β, and the objective costs of activities, c_1 and c_2.

To simplify the analysis, we assume that the aggregate benefit from prosocial activities and other public goods is linear[18] in A_1, A_2 and g, i.e.,

$$\int_0^1 V_i(A_1, A_2, g)d\theta_i = \gamma_1 A_1(\alpha) + \gamma_2 A_2(\alpha, \beta) + g, \quad (13)$$

where γ_1 and γ_2 are the relative weights the government puts on aggregate levels of activities 1 and 2 with respect to g. In order to obtain interior solutions, we assume $\gamma_1 \geq \gamma_2 > 0$. Finally, the aggregate income is simply the total amount of subsidies that citizens receive: $\int_0^1 I_i d\theta_i = s_1 A_1 + s_2 A_2$. Observe that this term does not appear in the expression of social welfare because subsidies are also included, with a negative sign, in the amount of public goods provided, g. Therefore, both terms cancel out. In fact, subsidies only affect social welfare through their influence on the supply of prosocial activities.

We plug the value $g = B - s_1 A_1 - s_2 A_2$ into function $W(s, g)$ in Eq. (10) and take into

[18] In Appendix C we show that all our results hold if we generalize the aggregate valuation of prosocial activities to a quasi-linear function $v(A_1, A_2) + g$, with $v_1 > 0$ and $v_2 > 0$.

account Eqs. (11) to (13) to express social welfare as

$$w(\alpha, \beta) = R + B + \gamma_1 A_1(\alpha) + \gamma_2 A_2(\alpha, \beta) - C(\alpha, \beta, c_1, c_2). \quad (14)$$

It is convenient to recall that α and β are the separating equilibrium thresholds characterized in Eqs. (3) and (4) and, as such, both depend on subsidies s_1 and s_2. The terms with positive sign in Eq. (14) are the benefits enjoyed by society from the government's provision of public goods and the private provision of prosocial activities. The term with negative sign represents the total private signaling costs associated with prosocial activities. A subsidy profile aimed at inducing maximization of social welfare must trade off these costs and benefits optimally.

In order to compute the optimal subsidies we proceed as follows: First, we compute the thresholds α^* and β^* that maximize $w(\alpha, \beta)$. Provided that the concavity of cdf $F(.)$ guarantees that function $w(\alpha, \beta)$ is strictly concave in both α and β, thresholds α^* and β^* are the solutions for α and β, respectively, of the equations system given by conditions $w_\alpha = 0$ and $w_\beta = 0$. Next, we evaluate functions $r_1(\alpha)$, $r_2(\alpha, \beta)$ and $r_0(\beta)$ in the optimal thresholds and write them as $r_1(\alpha^*)$, $r_2(\alpha^*, \beta^*)$ and $r_0(\beta^*)$. After that, we find the subsidies that induce these optimal thresholds. For this purpose, we solve for s_1 and s_2 the system formed by Eqs. (3) and (4), considering that reputations are given by $r_1(\alpha^*)$, $r_2(\alpha^*, \beta^*)$ and $r_0(\beta^*)$. The solution yields the optimal subsidies s_1^* and s_2^*.

The thresholds α^* and β^* that maximize function[19] $w(\alpha, \beta)$ are given by

$$\alpha^* = 1 - \frac{\gamma_1 - \gamma_2}{c_1 - c_2}, \quad (15)$$

$$\beta^* = 1 - \frac{\gamma_2}{c_2}. \quad (16)$$

[19] The maximizing thresholds are computed in the proof of Proposition 3, in Appendix A. In order to have $\alpha^* \geq \beta^*$ we must assume that the condition $c_1\gamma_2 \geq c_2\gamma_1$ holds.

The reason why optimal thresholds do not depend on the density function of the types is that this density affects the social benefits and the signaling costs of prosocial activities equally. The next result establishes the (optimal) subsidies that are successful in implementing thresholds α^* and β^*.

Proposition 3: *The optimal subsidy profile* $s^* = (s_1^*, s_2^*)$ *is characterized as*

$$s_1^* = \gamma_1 - [r_1(\alpha^*) - r_0(\beta^*)], \quad (17)$$

$$s_2^* = \gamma_2 - [r_2(\alpha^*, \beta^*) - r_0(\beta^*)]. \quad (18)$$

Proof: See Appendix A.

The expression of optimal subsidies s_1^* and s_2^* in Proposition 3 holds for any concave distribution of the agents's types[20]. These efficiency conditions have a fairly simple interpretation: the optimal subsidy on activity j is equal to the difference between the marginal social benefit from one more individual producing activity j and the marginal private benefit (reputation) that this individual can get by undertaking activity j instead of a_0. Therefore, optimal subsidies can be interpreted as a sort of Pigouvian subsidies in a scenario where production of prosocial activities generates positive externalities. At the optimal subsidy profile, the sum of both reputation and subsidy earned after contributing to activity j must be equal to the value for society of this contribution. Note that subsidies do not only internalize the (positive) externalities from the provision of social activities, but they also account for the reputation benefits individuals derive from contributing to prosocial activities. Therefore, the value of subsidies is, in general, lower than it would be in the benchmark case (i.e., without reputation signaling). It is worth mentioning that the characterization in Proposition 3

[20] Observe that, unlike optimal thresholds, optimal subsidies depend on the distribution of the individual's types.

also includes the case where subsidies may be negative (i.e., taxes). Basically, the sign of the right incentives depends upon the magnitude of the valuations of prosocial activities (see the example in the last section).

A distinctive feature of our approach is that individual reputation appears explicitly in the characterization of optimal subsidies. The public policy problem posed in this section can be summarized as follows. For any given array $(c_1, c_2, \gamma_1, \gamma_2)$ there exists an optimal partition of the interval $[0, 1]$ (given by the measure of agents that select each possible action) such that social welfare achieves a maximum. In order to find out which subsidy scheme manages to implement this partition (the optimal supply of each activity) the government must take into account that prosocial activities are used by the agents as an instrument to convey information about their willingness to cooperate with other agents. The agents' maximizing strategy together with the government's goal of maximizing social welfare shape the form of the right incentives for prosocial activities. Therefore, Eqs. (17) and (18) in Proposition 3 come from the combination of Eqs. (15) and (16) with Eqs. (3) and (4).

7 Comparative statics of optimal subsidies

In this section we explore how optimal subsidies respond to changes in their determinants, namely c_1, c_2, γ_1 and γ_2. We focus our attention on results that are apparently paradoxical, but whose rationale is grounded on the analysis of cross crowding out effects presented in Section 5. The precise formulae for the comparative statics results presented here are in Appendix A.

We first establish a useful result in order to understand the relationship between parameters c_1, c_2, γ_1 and γ_2 and the optimal thresholds α^* and β^* characterized in Eqs. (15)

and (16).

Lemma 3: $\frac{d\alpha^*}{dc_1} > 0$; $\frac{d\alpha^*}{dc_2} < 0$; $\frac{d\alpha^*}{d\gamma_1} < 0$; $\frac{d\alpha^*}{d\gamma_2} > 0$; $\frac{d\beta^*}{dc_1} = 0$; $\frac{d\beta^*}{dc_2} > 0$; $\frac{d\beta^*}{d\gamma_1} = 0$; $\frac{d\beta^*}{d\gamma_2} < 0$.

Proof: Straightforward from Eqs. (15) and (16).

Our first finding is that optimal subsidies vary asymmetrically in response to governmental valuations of prosocial activities, γ_1 and γ_2.

Corollary 1: $\frac{ds_2^*}{d\gamma_1} > 0$, and $\frac{ds_1^*}{d\gamma_2} < 0$.

Proof: See Appendix A.

We start by interpreting the sign of derivative $\frac{ds_2^*}{d\gamma_1} > 0$. Suppose that the social valuation of activity 1 increases. Then, from Lemma 3, threshold α^* must decrease while threshold β^* does not change. That is, the provision of activity 1 must increase but the aggregate supply of all activities must remain the same. This implies that the increase in activity 1 must be of the same magnitude as the decrease in activity 2. For this purpose, let us first consider raising s_1. Since $\alpha_1 < 0$, the supply of activity 1 is stimulated. However, this also brings about cross crowding out effects on the aggregate supply of prosocial activities, since $\beta_1 > 0$. Using the subsidy s_1 alone causes an inefficiently high reduction in the supply of activity 2. As long as $\beta_2 < 0$, the optimal response of the system also includes an increase in s_2, intended to mitigate excessive decline in the level of activity 2 caused by crowding out effects.

Let us now interpret the derivative $\frac{ds_1^*}{d\gamma_2} < 0$. Suppose that γ_2 increases. From Lemma 3, we have that α^* must increase and β^* must decrease. In other words, the optimal supply of activity 1 must be reduced and the optimal supply of activity 2 must be expanded. As in the previous case, the efficient way to induce these changes in thresholds comes from a certain combination of lower s_1 and higher s_2. Increasing s_2 may reduce the supply of activity 1 because $\alpha_2 > 0$, but at the same time it has a positive effect on the supply of all prosocial activities since $\beta_2 < 0$. Therefore, an increase in s_2 has the desired qualitative effects on activities 1 and 2. However, in order to achieve the exact value for the optimal supply of each activity, it is also necessary to reduce s_1. The reason is that the cross crowding out effect $\alpha_2 > 0$ is not strong enough, so a reduction in s_1 is required to reinforce the negative effect of increasing s_2 in the supply of activity 1.

Our second finding is that if activity 1 becomes more costly, it is optimal to reduce the subsidy on activity 2.

Corollary 2: $\frac{ds_2^*}{dc_1} < 0$.

Proof: See Appendix A.

The intuition for $\frac{ds_2^*}{dc_1} < 0$ is as follows. From Lemma 3, if c_1 increases (for instance), threshold α^* increases and threshold β^* remains constant. Since $\alpha_1 > 0$, the increase in α^* can be achieved through a lower s_1. However, as long as $\beta_1 > 0$, the lower subsidy s_1 induces a lower value of threshold β^*. Provided that $\beta_2 < 0$, the subsidy s_2 must decrease to keep β constant.

From the analysis above we conclude that governmental intervention in activities which may act as alternative devices for building reputation is far from being straightforward. Suppose that the government has an increasing concern about conservation activities (activity 1, in the example in the introduction). Nobody would doubt that s_1 should increase but few would claim that the efficient policy also includes increasing subsidies to volunteer caregiving (activity 2). However, according to our model, there exists a clear rationale for this combination of subsidies: Subsidizing conservation activities causes a decrease in the amount of volunteer caregiving, since some volunteers (the most cooperative ones) would change to conservation activities while others (the less cooperative ones) would abandon the activity. In order to mitigate cross crowding-out effects on activity 2, subsidy s_2 must also be increased. All the other paradoxical cases in the corollaries above can be interpreted in a similar way.

7.1 Example:

Consider the following example[21]: Individuals' types are uniformly distributed on the interval $[0,1]$, $c_1 = 2$ and $c_2 = 1$. Thus, thresholds α and β are given by $\alpha = \frac{4}{5} - \frac{4}{5}s_1 + \frac{2}{5}s_2$ and $\beta = \frac{3}{5} + \frac{2}{5}s_1 - \frac{6}{5}s_2$. In the context of the example, the condition $0 < \beta < \alpha < 1$ holds for a wide range of subsidies. For instance, if $s_1 = s_2 = 0$, the separating equilibrium is such that 60% of the population does not undertake any prosocial activity, 20% undertake activity 2 and 20% undertake activity 1. Suppose that $\gamma_1 = 2$ and $\gamma_2 = 1$. Is it socially beneficial to subsidize both activities? The answer is positive. Optimal subsidies in this case would be $s_1^* = \frac{3}{2}$ and $s_2^* = 1$. These subsidies would induce 100% of the population to undertake activity 1.

Let us check how this scheme responds to an increase in the cost of activity 1. Consider that $c_1' = 3$. Our theory predicts that optimal subsidies on both prosocial activities must decrease. To see this, we first use Eqs. (3) and (4) to compute the new separating equilibrium values for α and β as $\alpha = \frac{8}{9} - \frac{4}{9}s_1 + \frac{2}{9}s_2$ and $\beta = \frac{5}{9} + \frac{2}{9}s_1 - \frac{10}{9}s_2$. The new values for optimal subsidies are given by $s_1^{*\prime} = \frac{5}{4}$ and $s_2^{*\prime} = \frac{3}{4}$ (both have been lowered with respect to the values taken when $c_1 = 2$). According to Lemma 3, the optimal total amount of prosocial activities does not change. In fact, the new scheme induces half of the population to undertake activity

[21] The particular case of the uniform distribution considered in this example is developed in Appendix B.

1 and the other half to undertake activity 2. Finally, suppose that the valuation of activity 1 reduces up to $\gamma_1' = 1 = \gamma_2$. It is clear that in this case the optimal supply of activity 1 is zero. Subsidies on both activities now take the values $s_1^{*\prime\prime} = 0$ and $s_2^{*\prime\prime} = \frac{1}{2}$, and they induce 100% of the population to undertake activity 2.

It may also be the case that subsidies are negative (i.e., taxes). For instance, if the agents' types follow a uniform distribution, and $\gamma_1 = \gamma_2 = 1$, we have $s_1^* = \frac{1}{2} - \frac{1}{2c_2}$ and $s_2^* = \frac{1}{2}$. Clearly, $s_1^* < 0$ for $c_2 < 1$. If we reduce the social valuations of prosocial activities to $\gamma_1'' = \gamma_2'' = 0.5$ we have $s_1^* = -\frac{1}{2c_2} < 0$ for all c_2, and $s_2^* = 0$. In the extreme case where prosocial activities have null social valuations, it is optimal that all individuals choose not to undertake any activity. Thus, optimal taxes would be $s_1^* = -0.5$ and $s_2^* = -0.5$.

8 Conclusions

We have shown that the presence of competing signals for sending information about private characteristics linked to socially beneficial activities must be taken into account when designing policies based on rewards, incentives or fines on these activities. In settings where different activities are available, a subsidy or a fee imposed on one activity may have undesired effects if activities that act as potential substitutes are not considered. We study a context where agents signal their private characteristics through prosocial activities, and a public decision maker (who is well informed about the relevant signaling game) devises an incentive scheme that maximizes social welfare.

The formal approach adopted in this paper could also be applied (with minor changes to the model) to the analysis of career concerns problems and also to status games. For instance, in a career concerns model, the worker's output can be a signal of her ability.

The role of subsidies would be played by explicit incentives based on workers' observable performance. In this context, the compensation established on a given task may affect the number of workers accomplishing an alternative task due to cross crowding-out effects. In a status game, an agent's payoffs depend on others' perception of her wealth, which may be signaled through conspicuous consumption levels (actions a_0, a_1 and a_2). Wealthier people would face lower costs for each consumption level, and subsidies could be reinterpreted as tax reductions or exemptions established on certain goods. Again, as we proved in our model, cross crowding-out effects would play a relevant role in policy-making in this context. For example, if the goal of the policy-maker is to reduce the aggregate amount of conspicuous consumption, lowering taxes on consumption of the most luxurious goods (i.e. subsidizing action a_1) may be an efficient way of doing so.

In our model, prosocial behavior is driven by a blend of altruistic and self-interested motivations. Individuals have an instrumental concern for reputation, which is a purely private motivation. However, individuals are heterogeneous with respect to their willingness to cooperate, since they are endowed with a certain level of altruism. We may consider an extended model where individuals have preferences over prosocial activities. In this case, the agents' utilities should include the intrinsic motivation associated to the performance of each activity. We might not expect substantial changes in the modeling and conclusions if the preferences about activities depend on the agents' types. Regardless of the specific modeling of intrinsic motivation, including it in the analysis may hide a major idea in the paper. Namely, that cross crowding-out effects can appear when the relative costs of prosocial activities are modified by material incentives.

We use the insights derived from our signaling model to analyze the impact of subsidies on social welfare. Our conclusions go beyond the traditional analysis of optimal subsidization. The public decision maker must not only consider the budgetary cost of introducing subsidies compared to the social valuation of prosocial activities, but also the private signaling costs of prosocial activities and the individuals' reputation concerns. In this way, our approach embeds a classical problem of incentive design into a signaling structure in which individuals' reputations are critical in determining prosocial behavior. As shown in the paper, the existence of cross crowding-out effects supposes new complexities in the design of public policies.

References

[1] Ariely, D., Bracha, A. and Meier, S. (2009), Doing Good or Doing Well? Image Motivation and Monetary Incentives in Behaving Prosocially, *American Economic Review 99*, 544-55.

[2] Balakrishnan, N. (2001), Continuous Multivariate Distributions, John Wiley & Sons, Ltd.

[3] Bénabou, R. and Tirole, J. (2006), Incentives and Prosocial Behavior, *American Economic Review 96*, 1652-1678.

[4] Brekke, K.A., Kverndokk, S. and Nyborg, K. (2003), An Economic Model of Moral Motivation. *Journal of Public Economics 87*, 1967-1983.

[5] Bowles, S. and Hwang, S. H. (2008), Social Preferences and Public Economics: Mechanism Design when Social Preferences Depend on Incentives, *Journal of Public Economics 92*, 1811-1820.

[6] Bowles, S. and Polanía-Reyes, S. (2012), Economic Incentives and Social Preferences: Substitutes or Complements?, *Journal of Economic Literature 50*, 368-425.

[7] Buurman, M. and Dur, R. (2012), Incentives and the Sorting of Altruistic Agents into Street-Level Bureaucracies, *The Scandinavian Journal of Economics 114*, 1318-1345.

[8] Cho, I. K. and Kreps, D. M. (1987), Signaling Games and Stable Equilibria, *The Quarterly Journal of Economics 102*, 179-221.

[9] Dewatripont, M., Jewitt, I. and Tirole, J. (2000), Multitask Agency Problems: Focus and Task Clustering, *European Economic Review 44*, 869-877.

[10] Ek, C. (2017), Some Causes are More Equal than Others? The Effect of Similarity on Substitution in Charitable Giving, *Journal of Economic Behavior and Organization 136*, 45-62.

[11] Ek, C. (2018), Prosocial Behavior and Policy Spillovers: A Multi-Activity Approach, *Journal of Economic Behavior and Organization 149*, 356-371.

[12] Ellingsen, T. and Johannesson, M. (2011), Conspicuous Generosity, *Journal of Public Economics 95*, 1131-1143.

[13] Figenbaum, E., Kolbenstvedt, M. and Elvebakk, B. (2014). Electric Vehicles–Environmental, Economic and Practical Aspects, Institute of Transport Economics (TØI).

[14] Franke, T., Bühler, F., Cocron, P., Neumann, I. and Krems, J. (2012), Enhancing Sustainability of Electric Vehicles: A Field Study Approach to Understanding User

Acceptance and Behavior, in Sullman, M. and Dorn, L. (eds.), *Advances in Traffic Psychology*, Farnham.

[15] Frey, B. and Jegen, R. (2001), Motivation Crowding Theory: A Survey of Empirical Evidence, *Journal of Economic Surveys 15*, 589-611.

[16] Frey, B. and Oberholzer-Gee, F. (1996), The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out, *American Economic Review 87*, 746-755.

[17] Gneezy, U., Meier, S. and Rey-Biel, P. (2011), When and Why Incentives (Don't) Work to Modify Behavior, *Journal of Economic Perspectives 25*, 191-210.

[18] Gneezy, U. and Rustichini, A. (2000), A Fine is a Price, *Journal of Legal Studies 29*, 1-17.

[19] Greenwald, B. C. and Stiglitz, J. E. (1986), Externalities in Economies with Imperfect Information and Incomplete Markets, *The Quarterly Journal of Economics 101*, 229-264.

[20] Halvorsen, B. and Frøyen, Y. (2009), Trafikk i Kollektivfelt. Kapasitet og Avvikling. Elbilens Rolle. (Traffic in Public Lanes. Capacity and Handling. The Role of Electric Cars). Report from Asplan Viak.

[21] Holmstrom, B. and Milgrom, P. (1991), Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design, *Journal of Law, Economics and Organization 7*, 24-52.

[22] Holtsmark, B. and Skonhoft, A. (2014), The Norwegian Support and Subsidy Policy of Electric Cars. Should it Be Adopted by Other Countries?, *Environmental science and policy 42*, 160-168.

[23] Hong, S. (1999), The Effects of Unit Pricing System upon Household Solid Waste Management: The Korean Experience, *Journal of Environmental Management 57*, 1-10.

[24] Kinnaman, T. C. and Fullerton, D. (2000), Garbage and Recycling with Endogenous Local Policy, *Journal of Urban Economics 48*, 419-442.

[25] Kits, G. J., Adamowicz, W. L. and Boxall, P. C. (2014), Do Conservation Auctions Crowd Out Voluntary Environmentally Friendly Activities?, *Ecological Economics 105*, 118-123.

[26] Lilley, A. and Slonim, R. (2014), The Price of Warm Glow, *Journal of Public Economics 114*, 58-74.

[27] Rege, M. (2008), Why Do People Care about Social Status?, *Journal of Economic Behavior and Organization 66*, 233-242.

[28] Robèrt, M. and Jonsson, R. D. (2006), Assessment of Transport Policies toward Future Emission Targets: a Backcasting Approach for Stockholm 2030, *Journal of Environmental Assessment Policy and Management 8*, 451-478.

[29] Seabright, P. B. (2009), Continuous Preferences and Discontinuous Choices: How Altruists Respond to Incentives, *The BE Journal of Theoretical Economics 9.*

[30] Topkis, D. M. (1978), Minimizing a Submodular Function on a Lattice, *Operations Research 26*, 305-321.