How to turn crowding-out into crowding-in? An innovative instrument and some law-related examples

Antoine Beretti, Charles Figuieres, Gilles Grolleau

European Journal of Law and Economics
https://doi.org/10.1007/s10657-019-09630-9

How to turn crowding-out into crowding-in? An innovative instrument and some law-related examples

Antoine Beretti[1] · **Charles Figuières**[2] · **Gilles Grolleau**[3,4]

Abstract
Using a simple decision-theoretic approach, we formalize how agents with different kinds of intrinsic motivations react to the introduction of monetary incentives. We contend that empirical results supporting the existence of a crowding-out effect under various legal procedures hide a more complex reality, where some individuals contribute thanks to these additional monetary incentives while others reduce their contributions. Our approach allows us to study the theoretical ability of the *self selection mechanism* (Mellström and Johannesson in J Eur Econ Assoc 6:845–863, 2008; Beretti et al. in Kyklos 66(1):63–77, 2013) to reduce the likelihood to backfire against the cause it is meant to promote. This mechanism consists of a monetary payment for the pro-social behavior and it offers agents the choice to either keep the money for themselves or to direct it to a charity. We show that this legal procedure dominates others more classical procedures because it taps wisely into the motivational heterogeneity of individuals. It uses a self-selection mechanism to match adequate monetary incentives with individuals' types regarding intrinsic motivations. It may even turn a situation subject to crowding-out into a crowding-in outcome.

Keywords Crowding-out · Heterogeneity · Moral motivation · Environmental regulation

JEL Classification D03 · D64 · H23 · Q58

✉ Charles Figuières
Charles.Figuieres@univ-amu.fr

[1] Montpellier SupAgro, Montpellier, France

[2] CNRS, EHESS, Centrale Marseille, IRD, AMSE, Aix Marseille University, Marseille, France

[3] Burgundy School of Business-CEREN, Univesity Bourgogne Franche-Comté, Burgundy, France

[4] CEE-M, CNRS, INRA, Supagro, Université de Montpellier, Montpellier, France

Published online: 26 August 2019

1 Introduction

'Legal rules and regulations are routinely rationalized by appeal to the incentives they create', coined thereafter as legal incentives (Atiq 2014). Except some recent contributions (e.g., Feldman and Perez 2012; Atiq 2014; Underhill 2016), the complex way in which the law interacts with individuals' intrinsic and extrinsic motivations remains relatively neglected in the law and economics literature, even if some aspects have been tangentially discussed in the 'expressive law' literature, which holds that law influences behaviors independently of its sanctions (see for instance Cooter 2000). Our goal in this paper is twofold: (i) to provide a conceptual exploration of the subtle interactions between legal procedures and agents' motivations (intrinsic and extrinsic), (ii) to elaborate from it an adequate procedure—put more precisely, an incentive mechanism—in order to promote the desired behaviours.

Most existing contributions to date emphasize that legal incentives- e.g., incentives that are caused or created by laws and regulations such as jail sentences, monetary sanctions or compliance rewards—can crowd out the natural motivations of individuals—e.g., civic duty, moral or social motivations for compliance—to engage in socially desirable behaviours. Several scholars (Underhill 2016 and references therein) stressed that incentives-based legal rules can backfire in various contexts, such as paying organ donors or Good Samaritans, rewarding whistleblowers or compensating jurors. Atiq (2014) even argued that the desire to preserve the non-legal motives and the risk of crowding out these 'higher motives' constitutes a convincing explanation to the reluctance to implement legal incentives in some spheres of human conduct—marital relations, gift-giving and the scientific enterprise—that are governed by valued non-legal norms and motivations. For instance, in 2006, the famous Russian mathematician Grigory Perelman turned down the prestigious Fields Medal and 1 million-dollar-prize for proving the Poincaré Conjecture and explained that "everybody understood that if the proof is correct then no other recognition is needed." (Paulos 2010). Atiq (2014) also shows that taking into account the risk of crowding-out can give to standards a special advantage over bright-line rules. In his review, Bowles (2008) mentions four non-mutually exclusive mechanisms by which crowding-out occurs. These mechanisms can work jointly or sometimes have opposite effects. First, incentives may re-frame a decision problem and thereby suggest self-interest as the appropriate behavior. Second, incentives may affect the long-term development of preferences such as when introduction of fines for tardiness at day care centers induced more self-interested behavior, even after they were withdrawn. Third, incentives may 'overjustify' the activity and compromise the individual's sense of autonomy. Fourth, incentives convey information about the principal's preferences and beliefs concerning the agent and the nature of the task. For instance, just incentivizing people monetarily or monitoring them can make them interpreting the situation as if the principal does not trust them enough which in turn may lead them to exert less effort compared to the situation without monetary incentives.

The current view then holds that if monetary incentives—emphasized in the economic analysis—are a powerful instrument to change behavior, they are a part

of the story but *not the whole* story. Therefore, a crowding-out effect could lead to inferior outcomes. This would be due to an interaction, still poorly understood, between intrinsic motivation and extrinsic (dis)incentives introduced by the monetary instrument. The issue has been investigated, theoretically (Bénabou and Tirole 2003; Ellingsen and Johannesson 2008) and empirically (see Bowles 2008 for a recent review). It has been proven to be relevant in a wide variety of contexts such as blood donation (Titmuss 1970; Mellström and Johannesson 2008; Goette et al. 2010), acceptance of a polluting infrastructure (Frey and Oberholzer-Gee 1997) or to address late coming parents in day-care centers (Gneezy and Rustichini 2000) or bed-blocking in hospitals (Holmås et al. 2010).

In addition, most papers consider that *all* agents behave similarly when facing monetary incentives. Nevertheless, it is more realistic to assume that people are heterogeneous and have various intrinsic motivations according to the considered domain (Bénabou and Tirole 2006; Beretti et al. 2013). Feldman and Perez (2012) provide empirical evidence on how ignoring heterogeneity about intrinsic motivations among agents can misguide and undermine the efficacy of regulatory intervention. And this heterogeneity is probably the source of many difficulties encountered in defining the concept of intrinsic motivation, and controversies about its usefulness (Reiss 2005; Bruno 2013). Psychologists frequently consider that intrinsic motivations are those that arise from within—doing something because you want to—while extrinsic motivations mean people are seeking a reward, such as money or a trophy at a sporting event. Intrinsic motivation is that which is pleasurable *per se* (because it offers its own satisfactions), while extrinsic motivation is the desire to obtain something exogenous to the task. Put differently, one could resort to a means-end logic in order to determine whether a motivation is intrinsic or extrinsic: intrinsic motivation is doing what we want, whereas extrinsic motivation is doing something to get something else. Recently, Bolle and Otto (2010) proposed to define extrinsic motivations as added motivations that interfere and disturb a well-defined situation. This well-defined situation constitutes the reference where intrinsic motivations play their role. To measure intrinsic motivation, various approaches have been used such as self-reported measures of engagement and interest in the activity, observation of free-choice engagement in the activity when no rewards or other extrinsic motivators are present. Moreover, there are some studies investigating the possible neural basis of intrinsic motivation (e.g., Lee et al. 2012). They report differences in neural activation between intrinsic and extrinsic motivations.

We argue that any agent shelters a unique combination of various intrinsic motivations, and that this individual heterogeneity across agents is not without consequences on the effectiveness of public interventions. In particular, its overall impact in a whole population supporting or not the existence of a crowding-out effect can overlook the fact that various subgroups of this whole population react differently when facing the introduction of monetary incentives. With a focus on the overall impact of monetary incentives, in this paper we use the term 'crowding-out' ('crowding-in') to describe net behavioral outcomes, i.e., when the introduction of external incentives reduces (respectively, increases) the desired effort. This use is different from the common psychological use of the term which considers that crowding-out (crowding-in) occurs when intrinsic motivation is displaced (or

augmented) which then gives rise to a lesser (greater) engagement in the incentivized task. At the aggregate level, moderate monetary incentives can both motivate people who were originally not intrinsically motivated and harm the intrinsic motivations of people who were originally intrinsically motivated.

Notice that heterogeneity can refer to individuals who 'vary not only in level of motivation (i.e., how much motivation) but also in orientation of that motivation (i.e., what type of motivation) (Ryan and Deci 2000). Several studies reveal that incentive-based programs do not produce identical reactions across individuals (Gneezy et al. 2011) and psychologists recognize that different people are motivated in different ways (Reiss 2005; Lindenberg 2001). Even in the law literature, several scholars have made a case for taking into account the heterogeneity of intrinsic motivations among individuals (Feldman and Perez 2012; Underhill 2016). In the economic literature, Bar-Gill and Fershtman (2005) introduces heterogeneity by dividing people into two types and notably emphasize that public policies may influence norms and preference and consequently the distribution of these preferences among the population. These interdependences between policy and preferences may limit or enhance the effectiveness of different policies. Even if these authors do not explicitly consider legal incentives, we contend that their reasoning also applies to legal incentives that influence preference distribution among the population.

Natural candidates for heterogeneity in intrinsic motivations can be related to contextual parameters (Vohs et al. 2007) or other parameters such as intentions attributed to others or education. Even if they adopt a different viewpoint on intrinsic motivation, acknowledge individual differences in motivational orientation. Also, there is neurobiological basis for heterogeneity regarding altruism. Morishima et al. (2012) examined the neuroanatomical basis of individual differences in altruism with voxel-based morphometry and showed that gray matter (GM) volume in the right temporoparietal junction (TPJ) is strongly associated with both individuals' altruism and the individual-specific conditions under which this brain region is recruited during altruistic decision making.

We contend that integrating heterogeneity of intrinsic motivations can improve policymaking, by avoiding a one size-fits-all approach. Unfortunately most papers to date neglect this aspect and, in addition, do not propose alternatives to address the crowding out effect; they merely propose to eliminate monetary incentives. Two exceptions are, to our best knowledge, Mellström and Johannesson (2008) and Beretti et al. (2013), who suggest a mechanism that offers the agents the choice between a payment either for themselves or for a charity.[1] We refer to it as the *self-selection procedure*.

Our paper addresses these two issues in a theoretical model that (i) allows to take into account the heterogeneity of motivations—both in their nature and their degree—of individuals when faced with the introduction of monetary incentives (ii)

[1] It was an idea present in the air even before it has been studied in scientific articles. One of the authors of this article has once been offered a similar mechanism by a scientific journal, early in the 2000s. If his evaluation report was delivered in due time, he would be offered the choice of a payment either for himself or for an NGO caring for children in India.

studies the theoretical ability of Mellström and Johannesson (2008)'s mechanism to reduce the likelihood of getting a crowding-out result. We show that it achieves this property by tapping into the motivational heterogeneity of individuals. We assume that monetary incentives matter and change behaviors in predictable directions according to the matching between to whom they are directed (*i.e.* paying the individual versus paying the cause) and the preexisting level of intrinsic motivation of the individual (*i.e.*, low versus high level). Our model can explain a large variety of outcomes stressed in recent empirical studies and has policy relevance by suggesting a new instrument which eventually turns crowding-out into crowding-in (e.g., Beretti et al. 2013). Another interesting feature of this instrument lies in its ability to respect the principle of equality before the law while taking into account the heterogeneity of individuals. Indeed, heterogeneity is sometimes disregarded when designing legal incentives, because it frequently implies applying different policies to different people, which can offend ideas of justice and equal treatment. Although applying different policies to different segments of the population may maximize the effects of that policy, it may also invite backlash and threats to equality that undermine those benefits. Nevertheless, our proposed mechanism does not jeopardize the equality prerequisite, since all agents are faced with the same set of options, and it is they, not the regulator, who ultimately choose which option will be applied.

The rest of the paper is made of three sections. Section 2 develops a simple model that will serve as a conceptual tool for the exploration of the effect of several legal procedures. It culminates with the study of the self-selection procedure. Section 3 summarizes the main messages thanks to a simple example and discusses several law-related situations where considering heterogeneity is likely to influence law-making. Section 4 concludes.

2 Crowding-out with heterogenous agents: a simple model

This section constructs a simple behavioral model in which it is possible to explore the logical implications of various external monetary incentives on contributions when agents have heterogeneous intrinsic motivations. The model is framed in the environmental realm to fix ideas. For instance, we suggest that paying for recycling, or saving water, can push some people to reduce their recycling or saving behaviour if they were originally motivated by signaling their greenness through this means (see e.g., Thøgersen 2003). However, it is nevertheless clear that it could be applied to many domains where there is a mix of intrinsically motivated individuals and extrinsically motivated ones, such as volunteering or giving,[2] offering incentives for whistleblowing or rewarding jurors.

In the kind of situation we analyze in this paper, people know that they affect each other by their decisions, but their interactions are largely anonymous. They clearly don't know the set of strategies of the other individuals, nor do they know

[2] For instance, let us consider that intrinsically motivated donators are those who enjoy donating for its own sake. If we study experimental results of dictator games, we get a whole range of individuals from those giving nothing to those giving their own endowments.

their utility functions. Actually they even ignore how many "others" there are. Therefore we prefer to analyze the issue using a "decision-theoretic approach", in connection with the work of Bolle and Otto (2010), rather than with a "game-theoretic approach" as in B énabou and Tirole (2006).

There is a continuum of agents of unit mass. Each agent i is endowed with an exogenous income[3] y_i. The decision x_i of agent i is to contribute ($x_i = 1$) or not ($x_i = 0$) to some environmental cause and the opportunity cost of contributing, in monetary terms, is $c(x_i)$. The standard assumption is that this cost is an increasing function: $c(0) < c(1)$. Units are chosen in such a way that $c(0) = 0$, $c(1) = c > 0$. The remainder of the agent's income is affected to some alternative use $c_i = y_i - c(x_i)$.

The conceptual challenge of the present article is most entirely contained in the formulation of the objective function that the agents presumably maximize. Recall that we wish to capture heterogenous intrinsic motivations. And we want to give a role not only to the level of the motivation but also to its orientation. This last aspect in particular means that *procedural* consideration is an argument in the agents' objective functions, *i.e.* the same decision x_i performed under two different procedures can result in different perceived consequences by the agents. For evidence that people value not only outcomes, but also the procedures that lead to the outcomes see for instance Frey et al. (2004). Recognition of procedural concerns in agents' choice has recently led to reconsider both the field of *decision theory* (see Salant and Rubinstein 2008, for an axiomatic analysis of individual choices with frames) and that of *social choice* and *social welfare* (see Suzumura 1999; Suzumura and Xu 2001, 2003; Bernheim and Rangel 2007, 2009; Fleurbaey and Schokkaert 2013). And, inevitably, one must also reconsider the design of policy instruments, because they not only have an effect on x_i, but also because they are part of legal procedures, and legal procedures affect choices directly by themselves. This paper can be seen as a step in that direction.

Let f refer to the legal frame, or the procedure, that agent i faces in a particular choice situation. And assume that agent i is endowed with *decision-relevant* preferences, defined over bundles (x_i, f), numerically represented a *decision utility function* $U^i(x_i, f)$.We use here the popular distinction between *decision utility* which prompts actions, and *experienced utility,* or hedonic satisfaction,[4] which results from actions (see for instance Kahneman et al. 1997). It is fairly possible that the same decision x_i made under two different procedures f and f', produces different decision

[3] Agents have, possibly, very different exogenous incomes. Due to linearity assumptions in our model, this source of heterogeneity plays no role on agents' incentives. In a more complex model one can expect that wealthier (resp., poorer) people can be more interested by signaling their qualities (resp., gaining some money) than by the relative insignificant money incentive. The income dimension is also present in some legal incentives. For instance, in Finland speeding fines are linked to income, with penalties calculated on daily earnings. In 2015, a Finnish speeding millionaire was fined about 54.000 euros for being caught doing 103km/h in an area where the speed limit is 80km/h (https://www.bbc.com/news/blogs-news-from-elsewhere-31709454).

[4] Experienced utility functions would rather incorporate altruism, and capture the public good collectively created by inserting the others' aggregated contribution as an argument in utility functions of the kind $U^i(x_i, x_{-i}, f)$.

utilities: $U^i(x_i,f) \neq U^i(x_i,f')$. Put differently, to each procedure f corresponds a particular decision utility function $U^i_f(x_i) \equiv U^i(x_i,f)$.

We propose to study the effects of different procedures, with a focus on their interactions with intrinsic motivations. We consider four distinct procedures where all those ways are at play to different degrees: (N) a neutral procedure without external monetary incentives, (A) a procedure where individuals are paid for their contribution x_i, (B) a procedure where agents' decision to contribute is accompanied by a payment directed to a cause supporting the environment (say a relevant association or NGO), (C) a procedure where the agent is offered the choice regarding the orientation of the payment (for himself or for an association). The corresponding decision utility functions are denoted respectively $U^i_N(x_i)$, $U^i_A(x_i)$, $U^i_B(x_i)$, and $U^i_C(x_i)$. Below we make use of behavioral models, using specific functional forms for each $U^i_f(x_i)$, $f = N, A, B, C$. Beyond those functional forms, the interested reader is invited to check that generalizations of the results of the present paper are possible. The advantage of the simple forms we use is to offer a quick and clear way to highlight the logic we are studying.

2.1 Neutral procedure (N): pristine altruism alone

In the neutral procedure there is no incentive scheme and altruism is the only intrinsic motivation at work. Assume the decision utility function reads as:

$$U^i_N(x_i) = y_i - c(x_i) + a_i t^N x_i. \tag{1}$$

In expression (1):

- t^N is the marginal "monetarized" perceived benefits produced by the agent's contribution x_i on the other agents. This information parameter is procedure-dependent and it is associated here to the situation without external incentives;
- and $a_i \in [0, 1]$ is a parameter that captures an attitude towards the other individuals *via* the environment, a sort of ecologically-mediated, or 'green', altruistic concern.[5] Those parameters are uniformly distributed on $[0, 1]$ and each agent can be identified with a particular point in this interval. At one extreme, Agent 0 with $a_0 = 0$ does not feature any environmental concern; at the other polar case, Agent 1 with $a_1 = 1$ has a strong ecological conscience.

Assume that, for the most altruistic agent, with $a_1 = 1$, the marginal benefit of contributing covers its marginal cost, that is $c(1) - c(0) = c < t^N$. Individuals choose to contribute or not with a view to maximize (1). Hence, agents who settle for zero contributions are those such that:

$$y_i - c(0) > y_i + a_i t^N - c(1)\ ,$$

and the others contribute. Put differently, those in the interval $[0, a^N[$ where

$$a^N = [c(1) - c(0)]/t^N = c/t^N, \tag{2}$$

[5] The parameter a_i could also represent the degree of altruism or reciprocity.

are non-contributors, with a mass a^N, and those in the interval

$$C^N \equiv \left[a^N, 1\right]$$

are contributors. The total number of contributors is

$$1 - a^N. \tag{3}$$

2.2 Direct procedure (A): distorted altruism and moral repugnance

When individuals are paid for their contribution to the environment their decision utility function becomes:

$$U_A^i(x_i) = y_i - c(x_i) + wx_i + a_i t^A x_i - m\left(a_i, w, x_i\right), \tag{4}$$

where w is the monetary payment for participation. The introduction of the payment has two effects as far as intrinsic motivations are concerned.

First, the altruistic motivation is distorted. The idea is that the presence of a monetary transfer acts as a signal of the value of participation (Bolle and Otto 2010), upon which the agents' marginal benefits of the agents' altruism becomes $a_i t^A$ instead of $a_i t^N$, with $t^A < t^N$ because this parameter is procedure-dependent and the mere fact of paying agents for their altruism reduces its moral value compared to the procedure without payment. Moreover we assume that this price signal for altruism is at least equal to the payment offered ($t^A \geq w$).

Second, agents who have a concern for the environment now suffer from a *moral repugnance* associated with the fact of being automatically paid for contributing (see Roth 2007). This psychological aspect is often referred to - and studied - for another type of problem, that of organ donation (see Murray 1992). It is captured here by function $m(.,.,.)$ in expression (4). Putting a price onto a territory previously immune to the market forces is one of the list of events that generally spark the "yuck" factor argument (see for instance Kelly 2011; Sandel 2012, and also the discussion about obnoxious markets in Kanbur 2001). Clearly, moral repugnance is just one mechanism by which an incentivized person draws the conclusion that the principal offering the payment must have bad values. But there are lots of other mechanisms that would be impossible to distinguish from this. Here, moral repugnance stands in for all forms of crowding-out together. The monetarized value of this psychological "cost" is $m\left(a_i, w, x_i\right) \geq 0$. It is natural to assume that the larger the green altruism a_i, or the larger the payment w, and the larger the moral repugnance.[6] To put it formally, $m\left(a_i, w, x_i\right)$ is non decreasing in the two first arguments:

[6] In the case of organ donation, it has been argued that paying or more precisely compensating some expenses related to organ donation can support the donator's decision, but increasing the payment size by considering organs as a commodity activates a repugnance effect (Capron and Danovitch 2014; see also Roth 2007). In another context, the Resource Guide to U.S. Foreign Corrupt Practices Act published in November 2012 (http://www.justice.gov/sites/default/files/criminal-fraud/legacy/2015/01/16/guide.pdf) by the United States Department of Justice and the Securities and Exchange Commission stipulates regarding payments that "size can be telling, as a large payment is more suggestive of corrupt intent to influence a non-routine governmental action" (emphasis added).

$$\frac{\partial}{\partial a_i} m\left(a_i, w, x_i\right) \equiv m_a\left(a_i, w, x_i\right) \geq 0,$$
$$\frac{\partial}{\partial w} m\left(a_i, w, x_i\right) \equiv m_w\left(a_i, w, x_i\right) \geq 0.$$

We also assume that the marginal moral repugnance increases when altruism gets larger:

$$\frac{\partial^2}{\left(\partial a_i\right)^2} m\left(a_i, w, x_i\right) \equiv m_{aa}\left(a_i, w, x_i\right) \geq 0.$$

On the other hand, contributing could mitigate moral repugnance, so $m\left(a_i, w, x_i\right)$ is non increasing in the last argument:

$$m\left(a_i, w, 0\right) \geq m\left(a_i, w, 1\right).$$

Finally, without altruism, or without payment (when $w = 0$ or/and $x_i = 0$), there is no moral repugnance, therefore

$$m\left(0, w, x_i\right) = m\left(a_i, 0, x_i\right) = m\left(a_i, w, 0\right) = 0.$$

Agents decide to contribute or not by comparing the levels of utility attached to each possibility. Define the utility change of contributing:

$$\begin{aligned} \triangle U_A^i\left(a_i, w\right) \equiv & U_A^i(1) - U_A^i(0), \\ = & w + a_i\, t^A + c(0) - c(1) + m\left(a_i, w, 0\right) - m\left(a_i, w, 1\right), \\ = & w + a_i\, t^A - c - \triangle m\left(a_i, w\right), \end{aligned} \tag{5}$$

where $\triangle m\left(a_i, w\right) \equiv m\left(a_i, w, 1\right) - m\left(a_i, w, 0\right) = m\left(a_i, w, 1\right)$. Function $\triangle U_A^i\left(a_i, w\right)$ is supposed to be of class C^1 (meaning that $m(.,.,1)$ as a function of a_i and w is itself C^1). Notice that increasing the degree of green altruism can have two opposite effects on the utility change. The first effect is positive; it goes through the marginal benefit on others that is more valued by a more altruistic agent. The second effect is negative, because more altruism goes along with a more stringent moral repugnance under direct procedure A.

Contributors are those agents with $\triangle U_A^i\left(a_i, w\right) \geq 0$ and non contributors are agents i with $\triangle U_A^i\left(a_i\right) < 0$.

Assumption 1 Assume that:

$$\lim_{a_i \to 0} \triangle U_A^i\left(a_i, w\right) = w - c < 0\,, \tag{6}$$

$$\lim_{a_i \to 1} \triangle U_A^i\left(a_i, w\right) = w + t^A - c - \triangle m(1, w) < 0\,, \tag{7}$$

$$\lim_{a_i \to 0} \frac{d}{da_i} \triangle U_A^i(a_i, w) = t^A - m_a(0, w, 1) > 0\,, \tag{8}$$

$$\lim_{a_i \to 1} \frac{d}{da_i} \triangle U_A^i(a_i, w) = t^A - m_a(1, w, 1) < 0. \tag{9}$$

Item (6) of Assumption 1 focuses the analysis to payments w that are not high enough to encourage participation of the least altruistic agents (the payment alone is not enough to compensate the opportunity cost of contributing). This is the most interesting case, because if extrinsic incentives are too strong, no crowding-out effect can occur. Item (7) of Assumption 1 means that for the most altruistic agents, contributing is not optimal because their feeling of altruism towards others, though important, is overwhelmed by their moral repugnance of being paid. From items (8) and (9) of Assumption 1, given that:

$$\frac{d^2}{(da_i)^2} \triangle U_A^i(a_i, w) = -m_{aa}(a_i, w, 1) \leq 0,$$

and by the intermediate value theorem, $\exists a^*$such that $t^A - \frac{\partial}{\partial a_i} m(a^*, w, 1) = 0$. Therefore function $\triangle U_A^i(a_i)$ has an inverted U shape: it is first increasing, until a^*, then decreasing (see Fig. 1). Assume that $\triangle U_A^*(a^*) > 0$. Then there exists a neighborhood

$$C^A(a^*) \equiv \left[a^* - \underline{\varepsilon}, a^* + \overline{\varepsilon}\right] \sqsubset [0, 1]$$

of contributing agents around a^*, *i.e.* $\triangle U_A^i(a_i, w) \geq 0, \forall a_i \in C^A(a^*)$. Note that $C^A(a^*)$ is a *proper subset* of $[0, 1]$, for all the elements of $[0, 1]$ do not belong to $C^A(a^*)$; in particular, because of parts (6) and (7) of Assumption 1, agents a_0 and a_1 are not contributors. It is interesting to emphasize the peculiarity of this procedure. The choice to contribute can be explained by two intrinsic motivations of different natures: a degree of altruism sufficiently high or a moral repugnance not too strong (precisely in the most altruistic agents).

For future reference, let us denote:

$$\underline{a}^D = a^* - \underline{\varepsilon}\,,$$
$$\overline{a}^D = a^* + \overline{\varepsilon}.$$

the agents who, among those who contribute, have the lowest and largest altruism respectively. By definition, those two values solve the equation:

$$\triangle U_A^i(a_i, w) = w + a_i\, t^A - c - \triangle m(a_i, w) = 0. \tag{10}$$

Does crowding-out necessarily occur? To answer this question, one has to compare the mass of $1 - a^N$ of contributors under the neutral procedure with the mass $\overline{a}^D - \underline{a}^D$ of contributors under procedure A (see Fig. 1).

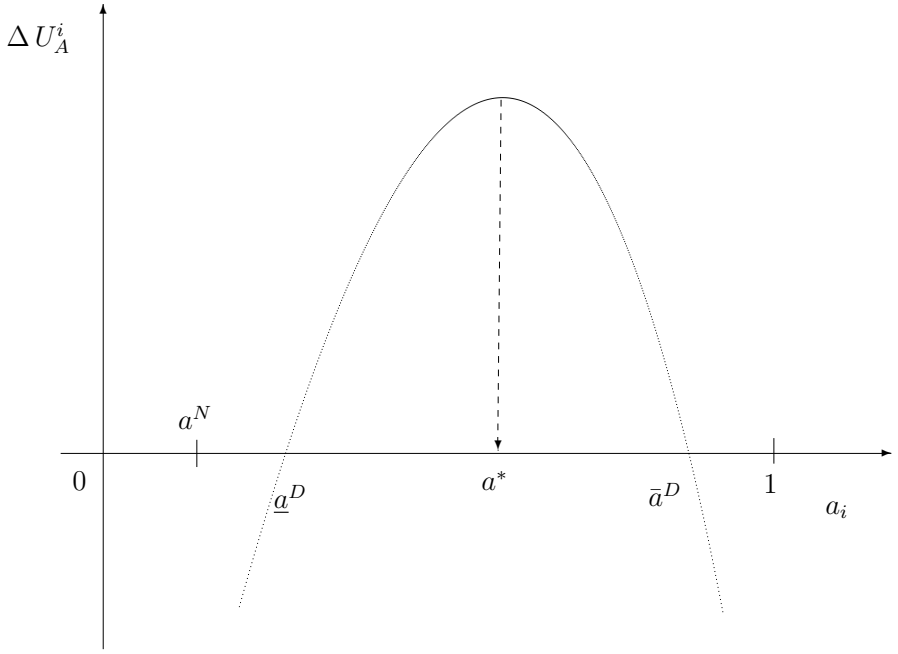

Fig. 1 Crowding-out occurs

Unless more structure is given to the moral repugnance function $\triangle m(a_i, w)$, Eq. (10) cannot be solved explicitly for $\underline{a}^D$ and $\overline{a}^D$. Still, important qualitative pieces of information can be obtained. The possibility of crowding-out depends on the relative position of a^N with respect to $\underline{a}^D$ and $\overline{a}^D$. The answer is ambiguous when $a^N \in \left[\underline{a}^D, \overline{a}^D\right]$ and when $a^N > \overline{a}^D$, but there is crowding-out for sure when $a^N \leq \underline{a}^D$ (Fig. 1 illustrates this case). All those situations can be associated with particular conditions on parameters. Since $\triangle U_A^i(a_i)$ has an inverted U shape and takes on positive values around a^*, by construction $a^N \in \left[\underline{a}^D, \overline{a}^D\right]$ if and only if:

$$\triangle U_A^i(a^N, w) \geq 0.$$

Using (5) and the fact that $a^N = c/t^N$ [see (2)]:

$$\triangle U_A^i(a^N, w) \geq 0 \quad \Leftrightarrow \quad w + \frac{c}{t^N} * t^A - c - m\left(\frac{c}{t^N}, w, 1\right) \geq 0. \tag{11}$$

By the same logic, if $a^N < \underline{a}^D$ or if $a^N > \overline{a}^D$, then necessarily:

$$\triangle U_A^i(a^N, w) < 0 \quad \Leftrightarrow \quad w + \frac{c}{t^N} * t^A - c - m\left(\frac{c}{t^N}, w, 1\right) < 0. \tag{12}$$

The last inequality means that, for agent a^N, the payment alone does not provide sufficient incentives to compensate the cost of moral repugnance and the decrease in altruistic motivation following the change of benefits on others from t^N to t^A. This is

consistent both with a too low value for w and with a too high value (recall that moral repugnance increases with w). This assessment of the weakness of monetary incentives is not absolute, but relative to parameters t^N and t^A. So, rewriting equality (12):

Definition ((*N/A*)-*weak extrinsic incentive*) Extrinsic incentives are (N/A)-weak when:

$$w < c\left(\frac{t^N - t^A}{t^N}\right) + m\left(\frac{c}{t^N}, w, 1\right).$$

The name (N/A)-weak is meant to express the property that the weakness of monetary incentive is relative to the erosion of altruism produced by the transition from procedure N to procedure A. The above reasoning has therefore established:

Proposition 1 *If* $a^N < \underline{a}^D$, *incentives are (N/A)-weak and there is crowding out.*

A last question is about the interactions between internal and external incentives. It is generally considered that the phenomenon of crowding-out gets weaker as external (monetary) incentives gets stronger. Let us analyze the effect of increasing the external incentive w on the upper and lower bounds of $C^A\left(a_i^*\right)$, *i.e.* on $\underline{a}^D$ and $\overline{a}^D$. By Eq. (10) and the implicit function theorem, using the properties that $t^A - m_a > 0$ until a_i^* and $t^A - m_a < 0$ after a_i^*, and under the assumption that moral repugnance increases less than proportionally with the external incentive, $m_w < 1$, we can conclude:

$$\left.\frac{da}{dw}\right|_{a=\underline{a}^D} = -\frac{1 - m_w\left(\underline{a}^D, w, 1\right)}{t^A - m_a\left(\underline{a}^D, w, 1\right)} < 0\,,$$

$$\left.\frac{da}{dw}\right|_{a=\overline{a}^D} = -\frac{1 - m_w\left(\overline{a}^D, w, 1\right)}{t^A - m_a\left(\overline{a}^D, w, 1\right)} > 0.$$

Therefore, as w increases the mass of contributors $C^A\left(a_i^*\right) = \left[\underline{a}^D, \overline{a}^D\right]$ gets wider. But it is important to notice that a crowding-out is not a necessary consequence of the direct procedure. A necessary condition for *crowding in* is $\underline{a}^D < a^N$. This may well happen if the monetary payment w is high enough. But this condition is not sufficient, because there is a mass of highly altruistic agents, $1 - \overline{a}^D$, who do not participate. By continuity, when w increases so that $\underline{a}^D$ decreases and falls below a^N, and if:

$$\left|\left.\frac{da}{dw}\right|_{a=\underline{a}^D}\right| < \left|\left.\frac{da}{dw}\right|_{a=\overline{a}^D}\right|,$$

that is if at the margin the increase of lower-end contributors is less than the decrease of upper-end non contributors, there is a continuum of values for w consistent with

crowding-in. Yet, the least altruistic agent who participates under the direct procedure is less altruistic than the least altruistic agent who participates in the neutral procedure.

To summarize,

Proposition 2 *Under Assumptions* (8), (9), *and when moral repugnance increases less than proportionally with the external incentive,* $m_w < 1$, *the stronger the external incentive w, the weaker the crowding-out effect, if any, under the direct procedure.*

2.3 Indirect procedure (B): distorted altruism alone

Under this design the payment is no longer given to contributors; rather it is directed to a cause supporting the environment, for example a related association or NGO (Bolle and Otto 2010). Hence individual i no longer bears the cost of moral repugnance, but of course his altruistic motivation is activated, for participation still generates a benefit to the environment.

The decision utility functions are now:

$$U_B^i(x_i) = y_i + a_i t^B x_i - c(x_i)\ ,$$

where t^B is the corrected marginal perceived benefits on the environment.

Regarding the fact that a payment is directed to the cause supported by the individual, we can reasonably consider that the perceived benefit on the environment of his participation is higher than when the same amount of money is directed to the individual's pocket, because the chosen destination, by its very nature, reinforces the belief of the agent on the presence of high environmental values, or because the association is more efficient than individuals in transforming a given amount of contributions in environmental gains. An assumption on parameters consistent with that view is $t^A \leq t^B$.

Hence, agents who settle for zero contributions are those such that:

$$y_i - c(0) > y_i + a_i\, t^B - c(1),$$

and the others contribute. Put differently, there is an interval $[0, a^I[$, where

$$a^I = \frac{c}{t^B}, \tag{13}$$

of non-contributors, and an interval

$$C^B \equiv \left[a^I, 1\right]$$

of contributors. The total number of contributors is

$$1 - a^I. \tag{14}$$

Compared to the neutral procedure there is crowding-out when $t^B \leq t^N$, because the mass of contributors has shrunk [compare expression (14) with expression (3)]. This non ambiguous result is rather intuitive: there is no direct own benefit and the estimation of the benefits on others has been cut down ($t^B \leq t^N$), so the incentives to participate are weaker compared to the neutral procedure.

However, the comparison with the direct procedure is more subtle. It follows a logic similar to the comparison of procedure A with procedure N. It is worth noting that there is a mass of agents near a_1 who contribute under the indirect procedure and who do not contribute under the direct procedure. In two cases, when $a^I \in \left[\underline{a}^D, \overline{a}^D\right]$ and when $a^I > \overline{a}^D$, we cannot state which policy better encourages participation. But the indirect procedure out-performs the direct procedure for sure when $a^I < \underline{a}^D$. Again all those situations can be associated with particular conditions on parameters. For $\underline{a}^D \leq a^I \leq \overline{a}^D$, a necessary and sufficient condition on the fundamentals of the model derives from the observation that $\triangle U_B^i\left(a^I, w\right) = \triangle U_B^i\left(c/t^B, w\right) \geq 0$ in such a situation. Or equivalently, using (5):

$$w \geq c * \frac{t^B - t^A}{t^B} + m\left(\frac{c}{t^B}, w, 1\right). \tag{15}$$

When this condition is not met, a necessary condition is obtained for $a^I < \underline{a}^D$:

Definition ((*B/A*)-*weak extrinsic incentives*) Extrinsic incentives are (B/A)-weak when:

$$w < c * \frac{t^B - t^A}{t^B} + m\left(\frac{c}{t^B}, w, 1\right).$$

When $a^I < \underline{a}^D$, incentives are (B/A)-weak and crowding-out is unambiguously less important under the indirect procedure. Intuitively, even if there are no monetary rewards for the agents under policy B, altruistic motives are less corroded than under the direct procedure and, in addition, the extrinsic motivation is not strong enough under policy A to compensate the weaker altruistic motivation and moral repugnance.

In a nutshell:

Proposition 3 *When the estimations of the benefits on others are such that* $t^A \leq t^B \leq t^N$, if $a^I < \underline{a}^D$ *the extrinsic incentives are (B/A)-weak and the indirect procedure unambiguously mitigates the crowding-out effect compared to the direct procedure. When extrinsic incentives are not (B/A)-weak, or when* $a^I > \overline{a}^D$*, the ability of the indirect procedure to improve participation compared to the direct procedure is ambiguous. However, in any case, participation under the indirect procedure is never larger than under the neutral procedure,* $a^N \leq a^I$.

2.4 Choice procedure (C): self-selection mechanism

Under this procedure, individuals can choose whether the payment is directed to themselves or to an environmental association. Giving the choice to individuals (keeping the reward for themselves or giving it to the 'environmental cause') could

motivate a wider set of individuals, possibly leading to a higher overall contribution.[7] At first glance, it can be surprising that receiving the payment incurs a cost of moral repugnance whereas directing a similar amount of money to the charity does not. From an economic perspective, one can argue that individuals can just redirect any payment that they receive to the charity they would like to support, thus turning any direct payment into a matching fund and avoiding the cost of moral repugnance. Nevertheless, psychological considerations can explain what seems irrational from a Homo œconomicus' viewpoint.[8]

In a sense, by choosing the target of the payment individual i chooses which decision utility function to activate.[9] Then, agent i's decision utility function U_C^i exists in two expressions:

- it is:

$$U_C^i(x_i) = U_A^i(x_i) = y_i - c(x_i) + wx_i + a_i t^A x_i - m\left(a_i, w, x_i\right) ,$$

  when the payment is direct.
- and it is:

$$U_C^i(x_i) = U_B^i(x_i) = y_i + a_i\, t^B x_i - c(x_i) ,$$

  when the payment is indirect.

Does the choice procedure minimize the countervailing effect of external incentives? Notice first that the utility attached to non participation is the same, whatever the chosen target:

$$U_C^i(x_i) = U_N^i(0) = U_A^i(0) = U_B^i(0) = y_i.$$

But the utility derived from participation differs according to the target of the payment. Agents who increase their utility by contributing are those who belong to at least one of the sets of contributors previously identified. Clearly, the set C^C of contributors under the choice procedure is the union of the two sets of contributors of each separate procedure, *i.e.* $C^C = C^B \cup C^A$. The set C^C encompasses agents a_i such that $\triangle U_B^i\left(a_i\right) \geq 0$, and/or $\triangle U_A^i\left(a_i\right) \geq 0$ and, therefore, the choice procedure

[7] Moreover, in addition to obvious intuitive reasons based on empirical evidence, we argue that people enjoy the possibility of choosing by themselves, even at a cost (Frey and Stutzer 2005).

[8] For instance, Tan and Low (2011) showed that subtle changes that do not make sense from an economic perspective change people's perceptions and behaviours. They explain that the words used to describe incentives for organs donors can dramatically change people's perceptions and subsequent behaviours. They cite the fact that the Singapore government paid a great deal of attention to the words used to describe these incentives (by avoiding the word 'payment' and preferring 'reimbursement' to 'defray the costs or expenses' associating with organ donation) in order to avoid crowding out prosocial motivations.

[9] We assume that offering choice does not affect parameters in utility functions. Nevertheless, a natural extension to our contribution will be to consider how the value of t and the shape of the moral repugnance function would be affected if individuals themselves can choose between the two options. Relaxing these these assumtions will considerably complicate the analysis and is beyond the scope of your analysis.

promotes participation as least as much as the two policies A and B separately do. But more precision can be added.

We will keep on assuming that estimations of the benefits of altruism are such that $t^A \leq t^B \leq t^N$. Even under this assumption,[10] several configurations for the different sets of contributors are possible:

Case 1 A first case is when $a^N \leq a^I < \underline{a}^D$, so the extrinsic motive is both (N/A)-weak and (B/A)-weak (definitions 2.2 and 2.3). Then the different sets of contributors are such that:

$$C^A \sqsubset C^B = C^C \sqsubseteq C^N.$$

Contributors under the choice procedure are exactly those who contribute under the indirect procedure and they are not more numerous than those who contribute under the neutral procedure.

Case 2 A more interesting case is when the monetary payment is sufficiently important to produce the following ranking $a^N \leq \underline{a}^D < a^I \leq \overline{a}^D$, that is the extrinsic motive is (N/A)-weak but it is not (B/A)-weak. The sets of contributors are then in the following configuration:

$$C^A, C^B \sqsubset C^C \sqsubseteq C^N.$$

Contributors under the choice procedure are more numerous than those who contribute under any of the two separate procedures. But the choice procedure does not perform any better than the neutral procedure.

Case 3 The most interesting case is when the monetary incentives are pushed slightly further so that $\underline{a}^D < a^N < a^I \leq \overline{a}^D$. The sets of contributors are such that:

$$C^A, C^B \sqsubset C^N \sqsubset C^C.$$

This is a case featuring crowding-out under each separate procedure, but there is crowding-in under the choice procedure. This possibility occurs because several intrinsic motivations exist and because agents are heterogenous. As a result those who contribute are not necessarily identical across procedures and, even more, C^A is neither a proper subset of C^B nor a proper subset of C^N. The corresponding necessary and sufficient conditions on parameters have been identified in (11) and (15). They must be imposed simultaneously, as a new assumption:

[10] Even if it is realistic, our assumption overlooks a possible information effect which appears when the payment offered acts as a signal that the cause is serious and worth-supporting.

Assumption 2 (Conditions for crowding-in)

$$w \geq c * \frac{t^N - t^A}{t^N} + m\left(\frac{c}{t^N}, w, 1\right) \text{ and} w \geq c * \frac{t^B - t^A}{t^B} + m\left(\frac{c}{t^B}, w, 1\right).$$

Case 4 Finally when $\underline{a}^D < a^N < \overline{a}^D \leq a^I$. It is not possible to conclude - without further information on the moral repugnance function - about the extent of the crowding-out phenomenon, if any, because there is a mass of agents characterized by intermediate degrees of altruism in the interval $]\overline{a}^D, a^I[$ who are not contributors. However, this situation is discarded when the extrinsic motivation is not t^B-weak.

To summarize, the choice procedure combines the incentive effects of both the direct and indirect procedures:

Proposition 4 *Let Assumption 1 holds and assume also $t^A < t^B \leq t^N$. Participation under policy C (choice procedure) is at least as large as under policies A and B. The choice procedure even results in crowding-in, although there is crowding-out under policies A and B, if and only if Assumption 2 is satisfied.*

3 Discussion

This brief section: (i) summarizes the messages of this paper *via* a simple example, (ii) develops several law-related examples where considering heterogeneity is likely to influence law-making.

3.1 Synthesis

Let us briefly use the practical example of a village exploiting a common resource, say a forest, in order to summarize and illustrate in words the expected outcomes of the four legal procedures. Under the neutral procedure N, there is no regulation. For cultural reasons, members of the village are informed of the level of exploitation that is socially optimal and intergenerationally fair. They partially comply to this collective norm, with some difference from one individual to another. They do so for a number of reasons: altruism, sense of duty, self-image, moral repugnance of over-destroying the forest and habitats, and so on. Under procedure A, a subsidy is introduced for reduced deforestation. Then the intrinsic motivations might be displaced or weakened, at least for a number of individuals: some of them realize the value they attributed to altruism was too high (compared to the value of the offered subsidy) and they lower it down, others feel their sense of autonomy and their moral image is compromised, and so on. A crowding-out outcome will be observed, unless the subsidy is sufficiently important compared to the erosion of moral motivations (see condition N/A-weak). Under procedure B, when the payment is directed to the cause, the counterproductive effect is mitigated compared to procedure A, though not eliminated. The "yuck" factor is less

stringent, but a price is still there to signal that altruism was over-rated (from the point of view of the regulator). Finally, under procedure C, the self-selection mechanism, society recognizes the plurality of motives and individuals are offered the choice of the society in which they want to live. Material interests and morals are not mixed up, their bad interactions are cut off.

A last remark is in order. Rather than just modifying the actual number of contributors in each scenario, there is also a kind of redistribution of roles (contributor or non-contributor). Behind the aggregated result, there are two simultaneous effects of crowding in and crowding out eventually counteracting each other. For example under procedure A, some highly motivated individuals may stop contributing whereas non-intrinsically motivated individuals may instead start contributing as long as the payments are high enough. The advantage of procedure C is to offer the best of both worlds.

3.2 Specific law-related examples

3.2.1 Whistleblowing

Some authors (Feldman and Lobel 2010) have argued that people vary in their intrinsic motivation level to report to authorities about misconduct. In a similar vein, Ayres and Braithwaite (1992) stated that a "strategy based mostly on punishment will undermine the good will of actors when they are motivated by a sense of responsibility." As a consequence, a population including people with low and high level of intrinsic motivation to obey the law by whistleblowing is likely to react differently to the same regulatory instrument. For instance, the findings of Feldman and Lobel (2010) "indicate that in some cases offering monetary rewards to whistleblowers will lead to less, rather than more, reporting of illegality" . Identifying the circumstances under which this outcome is likely and suggesting ways to avoid this counterproductive outcome is useful. We expect that adding and publicizing the possibility for the whistleblower to divert the payment to a credible fund against corruption or to the company's corporate social responsibility budget can lead to harness the best of both worlds, by retaining whistleblowing candidates that are intrinsically motivated without demotivating candidates who are more extrinsically motivated.

3.2.2 Tax compliance

Using data from a field experiment in a real-world context (mandatory church tax), Dwenger et al. (2016) found that the provision of compliance rewards has fundamentally different impacts on individuals with high level of intrinsic motivations (who increase their donations) and individuals with low level of intrinsic motivations (who increase their evasion). They assert: "That is, whether recognition for compliance raises or reduces tax payments hinges on what motivates taxpayers in the first place, with positive effects on the intrinsically motivated and negative effects on the extrinsically

motivated." Interestingly, some countries already use various compliance rewards to recompense honest taxpayers, such as Japan that offers the possibility to have your picture taken with the Emperor or the Philippines that put your name into a lottery (Feld et al. 2006).

Acknowledging and taking into account the heterogeneity of intrinsic motivation among tax payers allow to propose an adapted incentive mechanism. Rather than just rewarding monetarily and systematically good taxpayers, it is possible to design a set of compliance rewards that leaves to taxpayers the decision to direct or not the rewards to themselves or to another deserving cause, such as supporting a local community project.

3.2.3 Compensating jurors for participating in juries

The fact that jurors vary in their intrinsic motivation to participate is well-documented and the system of compensations of jurors is variable from a place to another with the possibility that employers still pay or not jurors from their companies. For instance, the US courts website (http://www.uscourts.gov/services-forms/jury-service/juror-pay) explains: "Federal jurors are paid $50 a day. While the majority of jury trials last less than a week, jurors can receive up to $60 a day after serving 10 days on a trial. (Employees of the federal government are paid their regular salary in lieu of this fee.) (...) Your employer may continue your salary during all or part of your jury service, but federal law does not require an employer to do so". For instance, Seamone (2002, p. 380) states that ' Among those jurors who are motivated by the sense of fulfilling a civic obligation rather than compensation, many may feel insulted by more than token compensation (emphasis added). To such people, substantial compensation detracts from the significance of their role as jurors and removes them from their temporary station of magistracy'.

Given that jurors compensations are frequently token and likely to cause (at least for some jurors) a kind of crowding-out, we argue that offering jurors the possibility to choose who will benefit from the rewards can decrease the likelihood of this outcome.

4 Conclusion

In line with Underhill (2016), we propose to harness the 'visible hand' of incentive architects to deliberately structure economic and legal incentives in order to address crowding-out effects. We propose that taking into account the possible crowding out of incentives at an early stage of law and regulation design can substantially increase their efficacy.

We made a strong case for how different intrinsic motivations among agents can play an instrumental role in explaining the effectiveness of introducing monetary incentives. We have formalized how the heterogeneity of intrinsic motivations among agents impacts their reactions to the introduction of monetary and other legal incentives. We showed that overall results supporting (or not) an undesired crowding-out effect can occult a more complex reality where some individuals contribute thanks to these additional monetary incentives while others reduce their contributions. Moreover, we studied a new instrument (Mellström and Johannesson 2008,

and Beretti et al. 2013) which taps into agents' heterogeneity in order to suppress, or at least to reduce, the risk of crowding-out result. This instrument avoids a 'one-size-fits-all' policy and allows agents to self-select the most relevant arrangement. A considerable advantage of this mechanism is that it does not require that policymakers and regulators have an extensive knowledge about the various levels of intrinsic motivations of agents.

A natural extension of our study is to examine the robustness of our main idea to various modifications of the setting. Is the result robust if we assume non-linearity both in intrinsic benefits and intrinsic costs (moral repugnance cost)? A host of evidence suggests that agents do not perfectly implement the optimal choices (or there is underlying heterogeneity along other dimensions than parameter a_i), what would happen if agents choose their optimal choices in proportions expressed by the logistic choice formulas, or when some other type of noise is introduced? What if agents' motivation depends on how many others contribute, or on the fraction of the population that contributes? What if agents care about the types of those whose welfare they promote, perhaps like to help those who are similar to themselves, or who are most pro-social? Experimental testing of these hypotheses could be also considered to update and improve the model. We also acknowledged that some legal incentives cannot be monetized and redistributed to charity, such as reducing someone's criminal sentence if he/she agrees to undergo drug treatment. These types of incentives are usually administered uniformly for reasons of equal protection and justice. Yet, in some cases, qualifications of the self-selection mechanism described in the paper could be contemplated, such as reducing someone's criminal sentence with time offered to charity organizations. In addition, it is fair to mention the recent appearance of some real-world programs that 'pay would-be criminals for good behaviours,[11] although these programs raise substantial ethical issues.

Theoretically, the proposed instrument respects the freedom of choice of individuals. Indeed, they decide about the final use of the received monetary incentives. Nevertheless, we are aware that the possibility of choice might also strongly vary with the framing of the task and could change the social behavior leading to different normative expectations. In line with the traditional maxim stressing that the evil is in the details, we encourage a careful design of real-world instruments by pretesting their various versions in pilot experiments.

References

Atiq, E. H. (2014). Why motives matter: Reframing the crowding out effect of legal incentives. *Yale Law Journal*, *123*(4), 1070–1117.

Ayres, I., & Braithwaite, J. (1992). *Responsive regulation: Transcending the deregulation debate*. New York: Oxford University Press.

[11] See https://www.independent.co.uk/news/world/americas/gun-crime-cash-rewards-therapy-holidays-richmond-california-operation-peacemaker-advance-peace-a7207601.html).

Bar-Gill, O., & Fershtman, C. (2005). Public policy with endogenous preferences. *Journal of Public Economic Theory*, *7*(5), 841–857.

Bénabou, R., & Tirole, J. (2003). Intrinsic and extrinsic motivation. *Review of Economic Studies*, *70*, 489–520.

Bénabou, R., & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, *96*, 1652–1678.

Beretti, A., Figuières, C., & Grolleau, G. (2013). Using money to motivate both saints and sinners: A field experiment on motivational crowding-out. *Kyklos*, *66*(1), 63–77.

Bernheim, D., & Rangel, A. (2007). Toward choice-theoretic foundations for behavioral welfare economics. *American Economic Review*, *97*(2), 464–470.

Bernheim, D., & Rangel, A. (2009). Beyond revealed preference: Choice theoretic foundations for behavioral welfare economics. *Quarterly Journal of Economics*, *124*(1), 51–104.

Bolle, F., & Otto, P. E. (2010). A price is a signal: On intrinsic motivation, crowding-out, and crowding-in. *Kyklos*, *63*, 9–22.

Bowles, S. (2008). Policies designed for self-interested citizens may undermine "The Moral Sentiments": Evidence from economic experiments. *Science*, *320*, 1605.

Bruno, B. (2013). Reconciling economics and psychology on intrinsic motivation. *Journal of Neuroscience, Psychology, and Economics, 6*(2), 136–149.

Capron, A. M., & Danovitch, G. (2014). We shouldn't treat kidneys as commodities, Los Angeles Times. Retrieved August 21, 2019, from http://www.latimes.com/opinion/op-ed/la-oe-0630-danovitch-and-20140630-story.html.

Cooter, R. (2000). Do good laws make good citizens: An economic analysis of internalized norms. *Virginia Law Review*, *86*, 1577–1601. https://doi.org/10.2307/1073825.

Dwenger, Nadja, Kleven, Henrik, Rasul, Imran, & Rincke, Johannes. (2016). Extrinsic and intrinsic motivations for tax compliance: Evidence from a field experiment in Germany. *American Economic Journal: Economic Policy*, *8*(3), 203–232.

Ellingsen, T., & Johannesson, M. (2008). Pride and prejudice: The human side of incentive theory. *The American Economic Review*, *98*(3), 990–1008.

Feld, L., Frey, B., & Torgler, B. (2006) Rewarding honest taxpayers? Evidence on the impact of rewards from field experiments. CREMA, Working Paper no. 2006-16.

Feldman, Y., & Lobel, O. (2010). The incentives matrix: The comparative effectiveness of rewards, liabilities, duties, and protections for reporting illegality (June 7, 2009). *Texas Law Review, 87*, San Diego Legal Studies Paper No. 09-013. https://doi.org/10.2139/ssrn.1415663.

Feldman, Y., & Perez, O. (2012). Motivating environmental action in a pluralistic regulatory environment: An experimental study of framing, crowding out, and institutional effects in the context of recycling policies. *Law and Society*, *46*(2), 405–442.

Fleurbaey, M., & Schokkaert, E. (2013). Behavioral welfare economics and redistribution. *American Economic Journal: Microeconomics*, *5*(3), 180–205.

Frey, B. S., & Oberholzer-Gee, F. (1997). The cost of price incentives: An empirical analysis of motivation crowding-out. *The American Economic Review*, *87*, 746–755.

Frey, B. S., Benz, M., & Stutzer, A. (2004). Introducing procedural utility: Not only what, but also how matters. *Journal of Institutional and Theoretical Economics*, *160*, 377–401.

Frey, B. S., & Stutzer, A. (2005). Beyond outcomes: Measuring procedural utility. *Oxford Economic Papers*, *57*, 90–111.

Gneezy, U., Meier, S., & Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives*, *25*(4), 191–210.

Gneezy, U., & Rustichini, A. (2000). A fine is a price. *The Journal of Legal Studies*, *29*(1), 1–17.

Goette, L., Stutzer, A., & Frey, B. M. (2010). Prosocial motivation and blood donations: A survey of the empirical literature. *Transfusion Medicine and Hemotherapy*, *37*(3), 481–502.

Holmås, T. H., Kjerstad, E., Lurås, H., & Straume, O. R. (2010). Does monetary punishment crowd out pro-social motivation? A natural experiment on hospital length of stays. *Journal of Economic Behavior and Organization*, *75*, 261–267.

Kahneman, D., Wakker, P., & Sarin, R. (1997). Back to Bentham? *Explorations of Experienced Utility, The Quarterly Journal of Economics*, *112*(2), 375–405.

Kanbur, R. (2001). On obnoxious markets. Revised version published. In S. Cullenberg & P. Pattanaik (Eds.), *Globalization, culture and the limits of the market: Essays in economics and philosophy*. Oxford: Oxford University Press (2004).

Kelly, D. (2011). *Yuck! the nature and moral significance of disgust*. Cambridge: MIT Press.

Lee, W., Reeve, J., Xue, Y., & Xiong, J. (2012). Neural differences between intrinsic reasons for doing versus extrinsic reasons for doing: An fMRI study. *Neuroscience Research*, *73*, 68–72.
Lindenberg, S. (2001). Intrinsic motivation in a new light. *Kyklos*, *54*(2/3), 317–352.
Mellström, C., & Johannesson, M. (2008). Crowding-out in blood donation: Was titmuss right? *Journal of the European Economic Association*, *6*, 845–863.
Morishima, Y., Schunk, D., Bruhin, A., Ruff, C. C., & Fehr, E. (2012). Linking brain structure and activation in temporoparietal junction to explain the neurobiology of human altruism. *Neuron*, *75*, 73–79.
Murray, T. H. (1992). The moral repugnance of rewarded gifting. *Transplantation & Immunology Letter*, *8*(1), 5–7.
Paulos, J. A. (2010). He conquered the conjecture, the new york review of books. Retrieved August 21, 2019, from http://www.nybooks.com/articles/2010/04/29/he-conquered-the-conjecture/.
Reiss, S. (2005). Extrinsic and intrinsic motivation at 30: Unresolved scientific issues. *The Behavior Analyst*, *28*, 1–14.
Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*, 54–67.
Roth, A. E. (2007). Repugnance as a constraint on markets. *Journal of Economic Perspectives*, *21*(3), 37–58.
Salant, Y., & Rubinstein, A. (2008). (A, f): Choice with frames. *Review of Economic Studies*, *75*(4), 1287–1296.
Sandel, M. J. (2012). *What money can't buy: The moral limits of markets*. Farrar, Straus and Giroux.
Seamone, E. R. (2002). A refreshing Jury Cola: Fulfilling the duty to compensate jurors adequately. *Legislation and Public Policy*, *5*, 289–417.
Suzumura, K. (1999). Consequences, opportunities, and procedures. *Social Choice and Welfare*, *16*(1), 17–40.
Suzumura, K., & Xu, Y. (2001). Characterizations of consequentialism and non consequentialism. *Journal of Economic Theory*, *101*, 423–436.
Suzumura, K., & Xu, Y. (2003). Consequences, opportunities, and generalized consequentialism and non-consequentialism. *Journal of Economic Theory*, *111*(2), 293–304.
Tan, C., & Low, D. (2011). Incentives, norms and public policy. In D. Low (Ed.), *Behavioural economics and policy design: Examples from Singapore* (Chap. 2, pp. 35–49). World Scientific Publishing Co.
Thøgersen, J. (2003). Monetary incentives and recycling: Behavioural and psychological reactions to a performance-dependent garbage fee. *Journal of Consumer Policy*, *26*, 197–228.
Titmuss, R. M. (1970). *The gift relationship: From human blood to social policy*. Sydney: Allen and Unwin.
Underhill, K. (2016). When extrinsic incentives displace intrinsic motivation: Designing legal carrots and sticks to confront the challenge of motivational crowding-out. *Yale Journal on Regulation*, *33*(1), 213–279.
Vohs, K., Mead, N. L., & Goode, M. R. (2007). The psychological consequences of money. *Science*, *314*, 1154–1156.