ENHANCING VOLUNTARY CONTRIBUTION IN A PUBLIC GOODS ECONOMY VIA A MINIMUM INDIVIDUAL CONTRIBUTION LEVEL

Documents de travail GREDEG
GREDEG Working Papers Series

Michela Chessa
Patrick Loiseau

GREDEG WP No. 2017-24
https://ideas.repec.org/s/gre/wpaper.html

Enhancing voluntary contribution in a public goods economy via a minimum individual contribution level

Michela Chessa[1] and Patrick Loiseau[2,3]

[1]Université Côte d'Azur, GREDEG, CNRS, 06560, Valbonne, France
[2]Inria, FairPlay team, Palaiseau, France
[3]Max-Planck Institute for Software Systems (MPI-SWS), Saarbrücken, Germany

GREDEG Working Paper No. 2017-24

Abstract

We consider the problem of increasing the total level of contribution in a public goods economy modeled as a non-linear public goods game. We propose a simple intervention that restricts the individuals strategy sets by imposing a minimum individual contribution level (while still allowing free-riding); and we investigate theoretically the potential of this intervention as a mechanism design tool. In particular we show that, for a well-chosen value of the minimum individual contribution level, our intervention does not incentivize any free-riding and strictly increases the total contribution level—compared to voluntary contributions—at a unique potential maximizer equilibrium. This is appealing because such a non-intrusive intervention is easily implementable in many different settings where normative or monetary interventions have met little acceptance. We present a motivating application in the domain of information economics about contribution of personal data to a data analytics project.

Keywords: Public goods, Voluntary contribution, Potential maximizer Nash equilibria, Minimum contribution level, Privacy

1 Introduction

In the last decades, the amount of personal information contributed by individuals to digital repositories such as social networks has grown substantially. The existence of this data offers unprecedented opportunities for data analytics research in various domains of societal importance, including public health, market research or political decision-making [Varian, 2014]. The results of these data analyses can often be considered as a *public good* that benefits data contributors as well as individuals who are not making their data available. At the same time, the release of personal information carries some privacy costs to the contributors: they may perceive the use of their data as an intrusion of their personal sphere [Altman, 1975, Warren and Brandeis, 1890] or as a violation of their dignity [Westin, 1970], or they may fear that this data can be abused, e.g., for unsolicited advertisements or social and economic discrimination [Acquisti and Fong, 2020, Mikians et al., 2013]. Indeed, evidence on data analytics projects shows a diffuse lack of individuals' confidence in information privacy [Malhotra et al., 2004], and as a consequence the presence of many individuals reluctant to disclose their data—the so-called *free-riders* in the classical public goods literature. In addition, many individuals may provide very small contributions, thus not being classified as free-riders, but not disclosing amounts of information significant to the data analytics research.

To avoid too small contributions, individuals taking part in data analytics studies are often asked to contribute at a minimum level. For example, for patients taking part in a medical study by filling out a survey, the answers to some questions are mandatory in order for the contribution to be taken into account in the study. More generally, a mechanism based on such a minimum contribution level is easy to implement in data science studies based on data held in an external database—which is often the case.[1] Indeed, individuals can decide on a "level of privacy" that they are comfortable with, or equivalently on a level of precision of the data they wish to contribute. Then, a local differential privacy mechanism is used to send the data in a way that ensures that the chosen level of privacy is met, which typically consists in adding noise to the data—with a well-chosen amount of variance—before sending it [Duchi et al., 2018]. It is then very easy to simply impose a minimum level of precision of the data contributed (or a maximum level of privacy). Still, participation remains voluntary and any individual can freely decide whether or not to take part in the study.

In such a mechanism, no one is pushed to participate, but the resulting contributions are naturally more substantial than in the original setting. Such an intervention is particularly appropriate in the domain of data release, in which enhancing higher contributions represents a delicate task. Indeed, it has been shown that individuals consider as fundamental to be able to exercise *control* over the release of their personal data [Damschroder et al., 2007, Kass et al., 2003, Robling et al., 2004] and they would not like an aggressive policy maker's intervention aiming at imposing their contribution. Moreover, monetary compensation has so far received very little traction and it has been shown to meet little acceptance in consumer surveys [Acquisti and Grossklags, 2005]. In contrast, letting individuals the freedom to participate or not, but asking for a minimum contribution in case of participation, represents a much more gentle intervention. Beyond the data analytics application, we observe similar settings in other more classical public goods models such as charitable giving, in which voluntary benefactors are usually asked to contribute a minimum amount (for example, the price of a charity raffle ticket). Yet, the effect and full potential of this intervention has not been analyzed systematically and deserves further investigation; this is our focus in this work.

In this paper, we investigate the potential of setting up a minimum individual contribution level as a mechanism design tool in public goods economies. We first propose a game-theoretic model that describes the basic interaction of individuals contributing to a public good as a non-linear public goods game where individuals choose their contribution level in the interval $[0, 1]$. Our model includes the case of data analytics projects as a special instance that models the trade-off individuals face when deciding on whether to share personal data—what Wieringa et al. [2021] call the *privacy calculus*.[2] In line with real-world evidence discussed above, our model captures the presence of very small contributions as a consequence of privacy concerns, which results in a sub-optimal total contribution to the public good. Then, we analyze how the introduction of a minimum contribution level can dampen this problem and how such a minimum contribution level can be optimally calibrated. Namely, we implement a variation of the aforementioned original public goods model, in which we restrict the strategy set of the individuals by imposing a minimum contribution level, and by still letting the possibility of choosing a zero contribution level. Restricting the strategy set is aimed at stimulating individuals toward higher efforts, while giving them the possibility of choosing a zero contribution ensures that such efforts remain voluntary. In particular, and as we will stress in the results of this paper, this intervention aims at increasing the contribution level of those individuals who are already willing to contribute in the original game, but with very small contributions. Instead, free-riders will remain coherent with their original choice and keep on free-riding. The main challenge in this intervention is that imposing a minimum contribution level can induce the adversarial side effect that some individuals previously

[1] In medical studies for instance, the data is often held in a database inside a hospital.

[2] Our model is in line with existing literature that already proposed to treat information as a public good [Abowd and Schmutte, 2019, McCain, 1988]. In some recent works, Ioannidis and Loiseau [2013], Gast et al. [2020], and Roussillon et al. [2022] (see also extensions of the initial model by Chessa et al. [2015a,b]) propose a different model which takes into account the public good nature of information. They analyze the game as a non-cooperative game and provide results on Nash equilibria and price of anarchy. An extension to a graph setting was proposed in Ramachandran and Chaintreau [2015]. The public good component of statistical analysis of data is also recognized by Abowd and Schmutte [2019]. In our model we assume continuous contributions, concave function of the total contributions and convex private costs. Convex private costs would arise, for instance, when the model represents situations in which the resources of an individual are allocated between public goods and private good consumption (as in [Bergstrom et al., 1986]) and they have been the object of alternative specifications of the classical public good model [see, e.g., Admati and Perry, 1991, Bramoullé and Kranton, 2007]. Moreover, we assume a classical *concave utility function* of the total contribution, which is mostly adopted in standard microeconomics models where the individuals are supposed to be risk-adverse and that characterizes a big part of the public good models literature as well.

contributing may now drop-off and become free-riders themselves, if they consider the minimum level to be too important. In this view, calibrating the optimal minimum contribution level is critical to the success of the intervention itself. It is important to note that the minimum contribution level can be interpreted as an individual threshold, but that our proposal fundamentally differs from researches on threshold public goods games [Cadsby and Maynes, 1999, Palfrey and Rosenthal, 1984, van de Kragt et al., 1983]. These studies investigate the impact of a minimum level for the total level of contributions or number of contributors in order for the public good to be provided, rather than a restriction of the strategy set (i.e., the individual contribution level), as is the focus of our work.

Our results show that it is possible to set up a minimum contribution level that improves the total contribution while reaching an equilibrium outcome where none of the individuals who would contribute in the original model have an incentive to free-ride as a consequence of the restriction of the strategy set. Specifically, there always exists a minimum contribution level such that, regardless of the possible multitude of Nash equilibria that may appear as a consequence of our intervention, there exists a unique potential maximizer equilibrium[3] in which all the contributors of the original game without minimum level are still contributing. Moreover, with such an equilibrium the total level of contribution is strictly larger than in the original game. Potential maximizer Nash equilibria, according to Monderer and Shapley [1996], are expected to accurately predict the results obtained through an experimental implementation of the model. Moreover, they are robust to incomplete information following the definition of Kajii and Morris [1997]. As such, they represent an important refinement of the equilibrium set in terms of applicability on real-world problems.

In the case of individuals with heterogeneous contribution costs (e.g., heterogeneous privacy concerns in the data analytics application), this minimum contribution level may not be expressed in closed form, but it can be easily computed from the game's parameters. We provide stronger results in case of homogeneous individuals. This simplified case is also realistic in practice as it corresponds to a case in which only the distribution of the individuals' contribution costs is known, hence one uses their expectation. In this simpler case, we show that for any minimum contribution level larger than the one we propose, some individuals who were previously contributing will start to free-ride. Under a mild condition (satisfied in particular in our data analytics example), this negatively affects the total contribution. In this homogeneous case, then, our minimum level of contribution is optimal.

In the context of data analytics projects, where privacy concerns are taken into account by a special instance of the public goods game with homogeneous individuals, it is possible to express our proposed minimum contribution level in closed form. Then our results show how a simple modification of the mechanism design allows a policy maker—here a data analyst—to substantially increase the accuracy of her analysis by correctly calibrating the minimum contribution level in the proposed survey (our simulations show a 25%-30% improvement on a simple population average estimation task, where the public good is given by the inverse variance of the corresponding estimation). Beyond this motivating example, however, we stress how interventions similar to the one we describe are widely applicable and easily implementable for a policy maker wishing to increase the provision of a public good above voluntary contribution levels in many other settings in which it is necessary to implement a non-intrusive intervention, and in which a policy maker intends to respect the choice of not contributing of free-riders. Our paper is intended to make this setting formal and to show the extent to which a mechanism (survey) designer can push forward this direction, by calibrating the minimum amount of required contribution (information) when the final goal is to increase the total level of contribution. Thus, to ensure the generality of our approach, we perform most of our analysis on a fairly general non-linear public goods model.

It is important to stress that the goal of our analysis—increasing of the total level of contribution—is often not aligned with the alternative objective of maximizing social welfare. This trade-off has already been discussed in works on public goods games about environmental cooperation, where the goal of increasing abatement levels favors environmental quality but results in so-called second-best agreements [Courtois and Haeringer, 2012]. In this paper—very similar in flavor to ours but profoundly different in the way the game is modified—a minimum abatement level is imposed in a second phase to players (countries) having decided strategically in a first phase to participate in an international environmental agreement, while non-

[3] Note that our public goods game is a potential game.

cooperating countries still have the original full set of strategies (including a zero abatement level). In the same flavor, our proposed solution can be interpreted as a second-best in terms of social welfare of the individuals participating in the public goods game.

The remainder of this paper is structured as follows. In Section 2, we present some literature on public goods economies, their inefficiencies and some ways to increase individuals' contribution. In Section 3, we present our generic theoretical model, and then the more specific model to describe a data analytics project as a public goods game. We briefly present in Section 4 the results on the status quo model without minimum contribution level that will be needed later. We then investigate the game with a minimum contribution level in Section 5, where we illustrate the potential of our approach. We discuss the results on the homogeneous case in Section 6, and we further detail such a case for the data analytics project in Section 7. We conclude in Section 8. All proofs are relegated to the Appendix.

2 Related work

Since the seminal work of Samuelson [1954], public goods are now omnipresent in the economics literature and model a broad range of goods, from fresh air to street lightning or national security. In many public goods settings, cooperation, even if individually costly, is sustained by the voluntary contributions of the involved individuals in a form of direct and indirect reciprocity [Rand and Nowak, 2013]. In other important instances, instead, tacit coordination as well as altruism or desire for fairness represent a sufficient gratification to enhance efficiency; for example in the context of private donations to charity [Roberts, 1984, Young, 1982], or environmental issues [Milinski et al., 2006]. In these cases, experimental analysis [see, e.g., McGinty and Milam, 2013] has shown the necessity to abandon the classical Nash equilibrium prediction in favor of an ethically-based rule of behavior [Sugden, 1982] in which preferences of people depend also on "*having done their bit*" [Cornes and Sandler, 1984]. We refer to Bergstrom et al. [1986], Kollock [1998] and Guido et al. [2019] for extensive reviews of theoretical studies and applications on the topic of social dilemmas and the voluntary provision of public goods.

However, as observed by Ledyard [1995], in many other settings "*one cannot rely on these approaches as a permanent organizing feature without expecting an eventual decline to self-interested behavior*". As a consequence of the non-excludability property, in fact, cooperation may lack or tend to decline over time [see, e.g., Burlando and Guala, 2005, Oliveira et al., 2015], letting room to the rise of equilibria that suffer from *free-riding* problems.

In such cases, reaching the objective of an efficient equilibrium outcome typically necessitates some monetary incentives or some normative interventions. Historically, various incentive compatible mechanisms focused on taxation. They enhanced efficiency in public goods economies [between the many, see, e.g., Clarke, 1971, Groves and Ledyard, 1977] based on the general framework for the mechanism design approach to welfare economics of Hurwicz [1972]. Since then, many other alternative mechanisms have been designed. Varian [1994] proposed a two-stage game in which individuals have the opportunity to subsidize the other individuals' contributions. In Andreoni and Bergstrom [1996], the idea was instead to let the government design a tax-financed subsidies scheme to increase the equilibrium contribution to a public good.

As alternative interventions not involving payments, it has been observed that cooperation dynamics may arise in the presence of *punishment* behaviors, i.e., when people who cooperate may be willing to punish free-riders. Punishment can have positive effects on the public good provision outcome. This has been shown to hold even when punishment is costly, and when no future benefits may be expected by the punishment activities. At this purpose, we cite the paper of Fehr and Gächter [2000] and all the follow-up studies [e.g., Herrmann et al., 2008]. However, it has been shown how the effect can break down in the presence of counter-punishment opportunities [Nikiforakis, 2008]. Another really productive line of research has instead investigated the effect of *reputation*, as need of social approval as consequent to one agent's action [Akerlof, 1980, Bénabou and Tirole, 2006]. Reputation has been stated to be essential for fostering social behavior among selfish agents, and it has been shown to be really effective when coupled to punishment [Sigmund et al., 2001]. Higher cooperation is also detected under meritocratic matching [Nax et al., 2014] or other kind of assortative matching of the agents, instead of random mixting: under assortative matching, agents contribute more than what would otherwise be strategically rational in order to be matched with others doing likewise [Duca et al., 2018]. Finally, we cite an original approach in the papers by Ledyard and Palfrey [1994, 2002], who proposed to approximate interim efficient mechanisms for the provision of a public good by a referendum in which voters simply vote for or against the provision of the public good.

However all these interventions do not fit well in many contexts in which monetary payments, punishment or all of the above approaches may be not natural or even "disturbing". In such domains, a policy maker may need to increase the provision of a public good above voluntary contribution less intrusively; ideally even without individuals being aware of the policy maker's intervention for encouraging contribution and enhancing efficiency. This is particularly true for our motivating example in the domain of information economics about voluntary contribution to data analytics projects [Acquisti and Grossklags, 2005].

3 The Model

3.1 The public goods economy

Our economy consists of a set $N = \{1, \ldots, n\}$ of individuals who may contribute to the provision of a public good. Each individual $i \in N$ has the same unit wealth and contributes to the public good at a (normalized) level $\lambda_i \in [0, 1]$. We denote by $\boldsymbol{\lambda} = [\lambda_i]_{i \in N} \in [0, 1]^n$ the vector of contributions and by $G(N) = \sum_{i \in N} \lambda_i \in [0, n]$ the total level of contribution of the individuals of set N.

In our model, we rely on voluntary contributions for the provision of the public good. For applicability reasons, we define a non linear public goods model with concave utility functions and convex cost functions[4]. Formally, given her contribution λ_i and the contributions of all the other individuals (for which we use the standard notation $\boldsymbol{\lambda}_{-i}$), i experiences a utility

$$U_i(\lambda_i, \boldsymbol{\lambda}_{-i}) = h(G(N)) - p_i(\lambda_i). \tag{1}$$

The first component, $h(G(N))$, is the *public good utility*, i.e., the homogeneous utility that each individual equally experiences because of the total provision of the public good. We suppose that $h : [0, n] \to \bar{\mathbb{R}}$[5] is twice continuously differentiable, that the individuals experience positive and diminishing marginal utility from the provision of the public good, hence h is strictly increasing and strictly concave. Moreover, we suppose that the marginal utility becomes negligible for high contribution levels, i.e., $h'(x) \to 0$ when $x \to +\infty$. The second component $p_i : [0, 1] \to \mathbb{R}_+$, represents the *cost of contribution* of individual i. Such a cost is heterogeneous, i.e., it depends on the type of each individual, but we suppose that all the p_i's are twice continuously differentiable, non-negative, increasing and strictly convex. Last, we suppose that for at least one $i \in N$, it holds that $h'(0) > p_i'(0)$. As we will show in the next section, this assumption ensures that, in our original model, at least one individual is voluntarily providing a strictly positive contribution. Analyzing public goods models in which the incentives to contribute are so low that everyone is free-riding is out of our scope, as we focus, instead, on improving an already existing non-zero total contribution. In particular, our aim is to induce non free-riders to contribute more.

Given the set of individuals N and a vector of contributions $\boldsymbol{\lambda} \in [0, 1]^n$, we define the associated *social welfare* as the function

$$W(\boldsymbol{\lambda}) = \sum_{i \in N} U_i(\boldsymbol{\lambda}).$$

Throughout our analysis and when specified, we often make the assumption that the individuals can be ordered in such a way that, for any contribution level $\lambda \in [0, 1]$, an individual choosing λ has higher marginal cost of contribution than the previous individuals if they choose the same contribution level. Formally:

Assumption 1. *The costs of contribution are such that $p_1'(\lambda) \leq \cdots \leq p_n'(\lambda)$, for all $\lambda \in [0, 1]$.*

Assumption 1 may require some re-ordering from the initial ordering.

[4]The assumptions of our model are well motivated by our application to a data analytics project that we present and discuss in Section 3.3.

[5]We denoted by $\bar{\mathbb{R}}$ the extended real number line $\mathbb{R} \cup \{-\infty, +\infty\}$.

3.2 The public goods game and the Nash equilibria

We represent the individuals' strategic interaction as the non-cooperative public goods game

$$\Gamma(N) = \langle N, [0,1]^n, (U_i)_{i\in N} \rangle ,$$

with set of players N, strategy space $[0,1]$ for each individual $i \in N$ and utility function U_i given by (1). We analyze the game $\Gamma(N)$ as a *complete information game* between the individuals, i.e., we assume that the set of individuals, the action sets and the utilities are common knowledge. A *Nash equilibrium* (NE) in pure strategies of this game is a strategy profile $\boldsymbol{\lambda}^* \in [0,1]^n$ satisfying

$$\lambda_i^* \in \arg\max_{\lambda_i \in [0,1]} U_i(\lambda_i, \boldsymbol{\lambda}^*_{-i}), \quad \forall i \in N.$$

Throughout our analysis, we always refer to NE (and the refinements) in pure strategies, even when not explicitly specified. A NE $\boldsymbol{\lambda}^*$ is a *strict Nash equilibrium* (SNE), if, for each $i \in N$ and for each $\lambda_i \in [0,1]$, $\lambda_i \neq \lambda_i^*$, it holds that

$$U_i(\lambda_i^*, \boldsymbol{\lambda}^*_{-i}) - U_i(\lambda_i, \boldsymbol{\lambda}^*_{-i}) > 0,$$

i.e., each individual strictly decreases her utility by deviating.

We observe that $\Gamma(N)$ is a potential game [Monderer and Shapley, 1996], with potential function $\Phi : [0,1]^n \to \bar{\mathbb{R}}$, s.t., for each $\boldsymbol{\lambda} \in [0,1]^n$,

$$\Phi(\boldsymbol{\lambda}) = h(G(N)) - \sum_{j\in N} p_j(\lambda_j). \tag{2}$$

In potential games, the set of global maxima is a subset of the Nash equilibrium set which refines it both in terms of equilibrium prediction and robustness. Indeed, first, in Monderer and Shapley [1996] the authors argue that such a refinement concept is expected to accurately predict the results obtained through an experimental implementation of the model; in particular, they show that this refinement concept accurately predicts the experimental results obtained by Van Huyck et al. [1990][6]. Second, this refinement concept has been justified theoretically in Oyama and Tercieux [2009] because of its robustness to incomplete information following the definition of Kajii and Morris [1997]. For these reasons, by implementing a potential game, we can assume that the individuals will converge to a *potential maximizer Nash equilibrium* (PMNE) $\boldsymbol{\lambda}^P \in [0,1]^n$ satisfying

$$\boldsymbol{\lambda}^P \in \arg\max_{\boldsymbol{\lambda}\in[0,1]^n} \Phi(\boldsymbol{\lambda}).$$

3.3 A data analytics project as a public goods game

Now, we model the interaction of the individuals in a data analytics project, and we underline how the game which arises is naturally embedded in the more general model we presented in the previous sections, that justifies our aforementioned hypotheses.

In a personal data economy, the set of individuals $N = \{1, \ldots, n\}$ may decide to voluntary contribute (or not) to a data analytics research project, providing some personal data at a given level of precision. We suppose that their personal data are collected and contained in a data repository. In particular, each individual $i \in N$ is associated with a *private variable* $y_i \in \mathbb{R}$, which contains sensitive information. We suppose that there exists $y_M \in \mathbb{R}$, s.t., the private variables are of the form

$$y_i = y_M + \epsilon_i, \quad \forall i \in N,$$

where ϵ_i are i.i.d., zero-mean random variables with finite variance $\sigma^2 < \infty$, which capture the inherent noise. We make no further assumptions on the noise; in particular, we do not assume that it is Gaussian.

[6]As we will discuss in the conclusions, an experimental validation of this affirmation on our model will be the main direction for future work on the topic.

As a result, such a model applies to a wide range of statistical inference problems, even cases where the distribution of variables is not known.

Parameter y_M represents the *mean* of the private variables y_i, and its knowledge is valuable to the analyst. The analyst wishes to observe the available private variables y_i and to compute their average as an estimation of y_M. In our model, we suppose that the analyst does not know the mean y_M that she wishes to estimate, but she knows the variance σ^2. Such an assumption is justified by the fact that in many statistical analyses observing the variability of an attribute in a population is easier than estimating the mean, both for the analyst and for the population (in Huberman et al. [2005], for example, the authors show how individuals value their age and weight information according to the relative variability).

We suppose that the analyst cannot directly access the private variables; instead she needs to ask the individuals for their consent to be able to retrieve the information. As such, the individuals have full control over their own private variables, and they have the choice to authorize or to deny the analyst's request. In particular, if wishing to contribute, but concerned about privacy, an individual can authorize the access to a perturbed value of the private variable. The *perturbed variable* has the form $\tilde{y}_i = y_i + z_i$, where z_i is a zero-mean random variable with variance σ_i^2 chosen by the individual. We assume that the $\{z_i\}_{i\in N}$ are independent and are also independent of the inherent noise variables $\{\epsilon_i\}_{i\in N}$. In practice, the individual chooses a given *precision* λ_i which is the inverse of the total variance (inherent noise plus artificially added noise) of the perturbed variable $\tilde{y}_i$, i.e.,

$$\lambda_i = 1/(\sigma^2 + \sigma_i^2) \in [0, 1/\sigma^2], \quad \forall i \in N,$$

and which corresponds to the contribution of individual i to the personal data economy. In the choice of the precision level, we have the following two extreme cases:

(i) when $\lambda_i = 0$, individual i has very high privacy concerns. This corresponds to adding noise of infinite variance or, equivalently, this represents the fact that individual i denies access to her data. In our public good model, a zero contribution corresponds to free-riding;

(ii) when $\lambda_i = 1/\sigma^2$, individual i has very low privacy concerns. This corresponds to authorizing access to the private variable y_i without adding any additional noise to the data. In our public good model, a maximal precision corresponds to a maximal level of contribution.

The strategy set $[0, 1/\sigma^2]$ contains all the possible choices for individual i: *denying*, *authorizing*, or any *intermediate level* of precision (which captures a wide range of privacy concerns as documented in behavioral studies [Spiekermann et al., 2001]). We denote by $\boldsymbol{\lambda} = [\lambda_i]_{i\in N}$ the vector of the precisions.

Once each individual $i \in N$ has made her choice about the level of precision λ_i and, consequently, the perturbed variable $\tilde{y}_i$ has been computed, the analyst has access to both the set of precisions and the set of perturbed variables. Then, the analyst estimates the mean as

$$\hat{y}_M(\boldsymbol{\lambda}) = \frac{\sum_{i\in N} \lambda_i \tilde{y}_i}{\sum_{i\in N} \lambda_i},$$

where perturbed variables with higher precision (i.e., smaller variance) receive a larger weight. This estimator is the standard *generalized least squares estimator*. It minimizes a weighted square error in which the i-th term is weighted by the precision of the perturbed variable $\tilde{y}_i$. This estimator is unbiased, i.e., $\mathbb{E}[\hat{y}_M] = y_M$, and has variance

$$\sigma_M^2(\boldsymbol{\lambda}) = \mathbb{E}[(\hat{y}_M(\boldsymbol{\lambda}) - y_M)^2] = \frac{1}{\sum_{i\in N} \lambda_i} \in [\sigma^2/n, +\infty]. \tag{3}$$

It is reasonable to assume that the analyst would use this estimator, as it is "good" for several reasons. In particular, it coincides with the *maximum-likelihood estimator* for Gaussian noise and, most importantly, it has minimal variance amongst the linear unbiased estimators for arbitrary noise distributions (this optimality result is known as Aitken theorem [Aitken, 1935]).

In the estimation, we have the following two extreme cases:

(i) when $\lambda_i = 0$ for each $i \in N$, the variance (3) is infinite. This corresponds to the situation in which each individual denies access to her data, and then the analyst cannot estimate y_M. In our public good model, this situation leads to a zero total level of contribution;

(ii) when $\lambda_i = 1/\sigma^2$ for each $i \in N$, the analyst estimates y_M with variance σ^2/n, resulting only from the inherent noise. This corresponds to the situation in which each individual is authorizing access to her data with maximum precision, i.e., no agent is perturbing her private variable. In our public good game, this corresponds to a maximal total level of contribution equal to n/σ^2.

For any level of precision in $[0, 1/\sigma^2]^n$, the estimated variance will be in $[\sigma^2/n, +\infty]$. The set of precision vectors for which the estimator has a finite variance is $[0, 1/\sigma^2]^n \setminus \{(0, \dots, 0)\}$.

We describe the interaction between the individuals as follow. We assume that each individual $i \in N$ wishes to minimize a cost function[7] $J_i : [0, 1/\sigma^2]^n \to \bar{\mathbb{R}}_+$, s.t., for each $\boldsymbol{\lambda} \in [0, 1/\sigma^2]^n$,

$$J_i(\lambda, \boldsymbol{\lambda}_{-i}) = c_i(\lambda_i) + f(\boldsymbol{\lambda}). \tag{4}$$

The cost function J_i of individual $i \in N$ comprises two non-negative components. The first component $c_i : [0, 1/\sigma^2] \to \mathbb{R}_+$ represents the privacy attitude of individual i, and we refer to it as the *privacy cost*: it is the (perceived or actual) cost that the individual incurs on account of the privacy violation sustained by revealing the private variable perturbed with a given precision. The second component $f : [0, 1/\sigma^2]^n \to \bar{\mathbb{R}}_+$ is the *estimation cost*, and we assume that it takes the form $f(\boldsymbol{\lambda}) = F(\sigma_M^2(\boldsymbol{\lambda}))$ where $F : [\sigma^2/n, +\infty) \to \mathbb{R}_+$ if the variance is finite, and $+\infty$ otherwise. It represents how well the analyst can estimate the mean y_M and it captures the idea that it is not only in the interest of the analyst, but also of the agents, that the analyst can determine an accurate estimate of the population average y_M. We assume that the privacy costs $c_i : [0, 1/\sigma^2] \to \mathbb{R}_+$, $i \in N$, are twice continuously differentiable, non-negative, non-decreasing, strictly convex and s.t. $c_i(0) = c_i'(0) = 0$, and that function $F : [\sigma^2/n, +\infty) \to \mathbb{R}_+$ is twice continuously differentiable, non-negative, non-decreasing and convex. To describe the strategic interaction between the individuals, we define the *estimation game* $\Gamma^E = \langle N, [0, 1/\sigma^2]^n, (J_i)_{i \in N} \rangle$ with set of agents N, strategy space $[0, 1/\sigma^2]$ for each agent $i \in N$ and cost function J_i given by (4).

We observe that the estimation game Γ^E is a particular case of the public good game $\Gamma(N)$ defined in the previous section. In fact, up to a normalization, or assuming $\sigma^2 = 1$, the strategy set of each individual corresponds to the normalized strategy set $[0, 1]$. Moreover, the minimization of the cost function J_i is equivalent to the maximization of a utility function $U_i^E : [0, n]^n \to \bar{\mathbb{R}}$ defined as

$$U_i^E(\lambda_i, \boldsymbol{\lambda}_{-i}) = h^E(\boldsymbol{\lambda}) - p_i^E(\lambda_i),$$

where the public good utility is given by $h^E(\boldsymbol{\lambda}) = -F\left(\frac{1}{\sum_{i \in N} \lambda_i}\right)$ and the cost of contribution by $p_i^E(\lambda_i) = c_i(\lambda_i)$. We may observe that this transformation leads to a negative utility. This is not in contradiction with the general formulation of the public good model we proposed, but it is, of course, unusual in the standard public goods literature. However, we observe that it would be sufficient to do a trivial re-scaling and translation of the utility function to write the same problem in a more ordinary form.

In the following, we present a detailed analysis for the more general model. Specific results about our application of a data analytics project are then presented in Section 7.

4 The Original Public Goods Economy

We may observe that the potential function in (2) is concave and the strategy sets are closed intervals of the real line. It follows that the set of Nash equilibria of $\Gamma(N)$ and the set of profiles which are maximizers of the potential function coincide. Then, exploiting its potential nature, we can perform a complete analysis of the

[7]We chose to present the estimation game model in its cost formulation for coherence with some previous work Chessa et al. [2015a,b]

game and provide its stable outcomes. In particular, we observe that a Nash equilibrium in pure strategies for such a game exists and it is unique, as stated in Theorem 4.1. As expected, at equilibrium individuals with higher contribution costs select lower contribution levels.

Theorem 4.1. *The game $\Gamma(N)$ has a unique NE $\boldsymbol{\lambda}^*$, which is strict and such that $\boldsymbol{\lambda}^* \neq (0, \dots, 0)$. If the costs of contribution satisfy Assumption 1, then the equilibrium is s.t., $0 \leq \lambda_n^* \leq \dots \leq \lambda_1^*$.*

In our model, owing to the assumption that $h'(0) > p_i'(0)$ for at least one $i \in N$, the equilibrium is such that we do not observe total free-riding behaviors by all the individuals. Remember that our aim is to induce non free-riders to contribute more. This result states that there exists a non-empty set of players that can be effectively pushed to contribute more. In the following, and whenever necessary, we use the notation $\boldsymbol{\lambda}^* = \boldsymbol{\lambda}^*(N)$ and $\lambda_i^* = \lambda_i^*(N)$ for each $i \in N$ to denote that the equilibrium depends on the specific identity of the agents in the set of individuals N (and, in particular, on their contribution costs).

Proposition 4.2 shows how, at equilibrium, the individual contribution and the total level of contribution vary when a new individual enters the game. In particular, we show that the individual contribution becomes lower, as expected and according to the standard public goods literature. However, the total level of contribution increases. The result holds without any ordering assumption, in particular, it holds irrespective of whether the new individual has higher or lower contribution cost compared to the rest of the population.

Proposition 4.2. *Given the game $\Gamma(N)$, suppose that an additional $(n+1)$-th individual enters the game. Then,*

(i) *the i-th individual's contribution at equilibrium for each $i \in N$ is s.t. $\lambda_i^*(N \cup \{n+1\}) \leq \lambda_i^*(N)$;*

(ii) *the total level of contribution is s.t. $G^*(N \cup \{n+1\}) \geq G^*(N)$.*

As a direct consequence of Proposition 4.2, we observe that, at equilibrium, both the individual utility and the social welfare are bigger when a new individual enters the game. These results are presented in Corollary 4.3.

Corollary 4.3. *Given the game $\Gamma(N)$, suppose that an additional $(n+1)$-th individual enters the game. Then, at equilibrium, $U_i(\boldsymbol{\lambda}^*(N \cup \{n+1\})) \geq U_i(\boldsymbol{\lambda}^*(N))$ for all $i \in N$, and $W(\boldsymbol{\lambda}^*(N \cup \{n+1\}))) \geq W(\boldsymbol{\lambda}^*(N))$.*

It is worth noticing that the inequalities of Proposition 4.2-*(ii)* and of Corollary 4.3 hold strictly under some specific assumptions; for instance, whenever the new individual is such that $h'(G^*(N \cup \{n+1\})) > p_{n+1}(\lambda_{n+1}^*(N))$ (i.e., if she is not a free-rider once she enters the game) and, in particular, whenever Assumption 1 holds and $n+1$ is ranked in the ordering no higher than the last non free-rider of the original game $\Gamma(N)$[8].

The results of this section, which are valid for a fairly classic non linear public goods model, adapt to our specific context some well known results already present in the literature [see, e.g., Bergstrom et al., 1986]. We can summarize the results presented in this section by affirming that our economy, modeled by the public goods game $\Gamma(N)$ and when we rely on the individuals' voluntary contribution, is inefficient in terms of total level of contribution. However, enlarging the set of potential contributors to the public good may be beneficial at this scope. Moreover, also from both the individual and the social welfare point of view, it is convenient to welcome as many individuals as possible into the economy. In a real-world application, this translates into proposing contributing to the public good to as many individuals as possible. However, improving the equilibrium provision level by enlarging the set of individuals is often not feasible, as some objective constraints could prevent this solution. This is mainly given by the fact that potential contributors are *a priori* limited (e.g., the set of users of a given platform is finite). In the next section, Section 5, we propose and illustrate a new and alternative way to increase the total contribution level, while maintaining the population of potential contributors fixed.

Some additional results on the original public goods economy under the more restrictive assumption of homogeneous contribution costs are presented in Section 6.1.

[8]This second condition is sufficient, but not necessary for the inequalities to hold strictly.

5 The Modified Public Goods Economy

5.1 The modified public goods game

As we have seen in the previous section, the public goods game $\Gamma(N)$, representing our economy and relying on the voluntary contribution of the individuals, has a unique Nash equilibrium. In such a model, the resulting total level of contribution, $G^*(N)$, can be inefficiently low. In the following, we suppose that the equilibrium contribution of $\Gamma(N)$ is such that $\lambda_i^*(N) < 1$ for some $i \in N$, i.e., it is such that the total level of contribution is not maximal, and then there is room for improvement.

In this section, we propose a way to enhance a higher level of contribution by the individuals. The resulting game is a variation of the original one proposed in the previous section. It is simply based on a restriction of the individuals' strategy space to $\{0\} \cup [\eta, 1]$ for a given $\eta \in [0, 1]$, that we call the *minimum contribution level*. Restricting the strategy set can stimulate the individuals toward higher efforts; letting them the possibility of choosing a zero level of contribution, we ensure that such efforts remain voluntary. As a drawback, some individuals may decide to stop contributing to the public good, and make our economy even more inefficient. We investigate how to tune the minimum precision level in order to improve the total level of contribution.

To analyze the strategic interaction between the agents in this variation, we define the game $\Gamma(N, \eta) = \langle N, \left[\{0\} \cup [\eta, 1]\right]^n, (U_i)_{i \in N} \rangle$, where the utility function U_i is still defined by (1), but it is now restricted on the domain $\left[\{0\} \cup [\eta, 1]\right]^n$. Observe that the original game $\Gamma(N)$ is a special case of this modified game $\Gamma(N, \eta)$, when $\eta = 0$. Then, from now on, we suppose that $\eta \in (0, 1]$. As we did for $\Gamma(N)$, we analyze the game $\Gamma(N, \eta)$ as a *complete information game* between the individuals, i.e., we assume that the set of individuals, the action sets (in particular, the minimum contribution level η) and the utilities are common knowledge. A Nash equilibrium (in pure strategy) of the modified game $\Gamma(N, \eta)$ is a strategy profile $\boldsymbol{\lambda}^*(\eta) \in \left[\{0\} \cup [\eta, 1]\right]^n$ satisfying

$$\lambda_i^*(\eta) \in \underset{\lambda_i \in \{0\} \cup [\eta, 1]}{\arg\max} \; U_i(\lambda_i, \boldsymbol{\lambda}_{-i}^*(\eta)), \quad \forall i \in N.$$

5.2 The Nash equilibria of the modified public goods game

We observe that the modified game $\Gamma(N, \eta)$ is still a potential game, with potential function Φ as in (2), but defined on the restricted domain $\left[\{0\} \cup [\eta, 1]\right]^n$. Differently from $\Gamma(N)$, the strategy sets of the modified game are not closed intervals of the real line and then the set of global maxima of the potential function may not coincide with the set of Nash equilibria (recall, though, that it still represents a refinement of it). As a consequence, we can no longer exploit directly its potential nature to perform a complete analysis of the Nash equilibrium structure, as we did in the previous section. However, we observe that, given its strict concavity, the potential function Φ has still a unique global maximum when restricted to the smaller and convex domain $[\eta, 1]^n$. We denote by $\boldsymbol{\lambda}^M(N, \eta) \in [\eta, 1]^n$ such a maximum

$$\boldsymbol{\lambda}^M(N, \eta) = \underset{\boldsymbol{\lambda} \in [\eta, 1]^n}{\arg\max} \, \Phi(\boldsymbol{\lambda}),$$

and by $G^M(N, \eta)$ the corresponding total level of contribution. In the following, we show how these quantities still play a fundamental role in the analysis of the Nash equilibrium structure of the game $\Gamma(N, \eta)$.

At first, we observe that introducing a minimum level of contribution, we may increase the number of free-riders. Some individuals (in particular, the ones with the highest contribution costs) may now find more advantageous to stay out of the game (i.e., contribute zero) rather than being pushed to contribute more. From now on we suppose that the game $\Gamma(N)$ is such that there are no free-riders. Formally:

Assumption 2. *The game $\Gamma(N)$ is such that at the unique NE $\boldsymbol{\lambda}^*$, $\lambda_i^* > 0$ for each $i \in N$.*

Under this assumption, ensuring that the number of free-riders does not increase translates into ensuring that in the modified game there are still no free-riders.

We observe that when we restrict our analysis to games which satisfy this assumption, we make our study more straightforward, but we do not lose any generality. In fact, for any game $\Gamma(N)$, we may consider an equivalent game reduced to the non free-riders, with utility function still defined as in (1), but up to an additive constant[9]. All our findings for such reduced game may be generalized to the original one, simply observing that when imposing a minimum contribution level, the free-riders will keep on free-riding and our incentives can be calibrated on and intended for increasing the effort of individuals who were already contributors to the public good.

Given Assumption 2, in Definition 5.1, we introduce a parameter η^* such that, as we see in Proposition 5.2, if we set a minimum precision level in $(0, \eta^*]$, there still exists a NE such that even the individual with the highest cost of contribution does not have an incentive to drop out and to free-ride and such an equilibrium coincides with $\boldsymbol{\lambda}^M(N, \eta)$.

Definition 5.1. *Suppose that Assumptions 1 and 2 are satisfied. We define the parameter η^* as the smallest parameter in $(\lambda_n^*, 1]$,[10] if there exists one, s.t.,*

$$h(G^M(N, \eta^*)) - h(G^M(N, \eta^*) - \eta^*) = p_n(\eta^*). \tag{5}$$

If such a parameter does not exist, we set $\eta^* = 1$.

To better understand this definition, we first observe that, when choosing a minimum contribution level $\eta \in (\lambda_n^*, 1]$, vector $\boldsymbol{\lambda}^M(N, \eta)$ is necessarily on the border (as no critical point of Φ was in $(\lambda_n^*, 1]^n$), i.e., it is s.t. the individual with the highest cost contributes exactly $\lambda_n^M(N, \eta) = \eta$. Given that, according to Theorem 4.1, the equilibrium of the original game is strict, intuitively, we define η^* as the smallest contribution level at which, if there still exists an equilibrium in which all the individuals are contributing, such an equilibrium is not strict anymore, i.e., individual n has exactly the same utility by contributing and by free-riding. If such a parameter does not exist, it follows that individual n is always better off by contributing than by free-riding, regardless of the minimum contribution level, and then we set it at maximum, i.e., equal to 1. In that case, it simply holds that

$$h(G^*(N, 1)) - h(G^*(N, 1) - 1) = h(n) - h(n-1) > p_n(1). \tag{6}$$

We now investigate whether, by introducing a minimum precision level, we still have an equilibrium in which also the individual with the highest contribution cost is not free-riding. In Proposition 5.2 we show that, by setting a precision level no bigger than η^*, this is still possible, and moreover such an equilibrium is the unique one with non-zero components.

Proposition 5.2. *Suppose that Assumptions 1 and 2 are satisfied. It holds that:*

(i) *if $|N| = 1$, then for any $\eta \in (0, \eta^*)$, $\Gamma(N, \eta)$ has a unique NE $\lambda^*(N, \eta) = \max\{\lambda^*, \eta\}$, where λ^* is the single entrance NE of $\Gamma(N)$ when only one individual is present. When $\eta = \eta^*$, there are two NE in 0 and in η^*, while for any $\eta \in (\eta^*, 1]$, there is a unique NE in 0;*

(ii) *when $|N| > 1$, then for any $\eta \in (0, \eta^*]$, every NE $\boldsymbol{\lambda}^*$ of $\Gamma(N, \eta)$ is s.t. $\lambda_i^* = \lambda_i^M(S, \eta)$ for each $i \in S$, where $S = \{i \in N | \lambda_i^* \neq 0\}$. In particular, $\boldsymbol{\lambda}^*(N, \eta) := \boldsymbol{\lambda}^M(N, \eta)$ is the unique NE of $\Gamma(N, \eta)$ s.t. $\lambda_i^*(N, \eta) > 0$ for all $i \in N$.*

Proposition 5.2 translates the fact that, if we select a minimum precision level that is not "too high", i.e., that is not bigger than η^*, there exists an equilibrium where all the individuals are still willing to contribute. This does not exclude the existence of other equilibria where some individuals may contribute zero. Whether the result is still valid or not for value of η bigger than this threshold is still an open question, but we have seen that this is not true for the special case of $|N| = 1$. However, as better stated in the final part of this section, we conjecture that this result will not hold anymore for a minimum contribution level which is larger

[9] The additive constant is given by the sum over i of the $p_i(0)$ with i previous free-rider.

[10] Recall that λ_n^* denotes the equilibrium contribution level in $\Gamma(N)$ of the individual with the highest contribution cost.

than η^* even for a larger set of individuals, translating the fact that, in that case, the individuals would feel to be forced to a too important effort and prefer not to cooperate (e.g., not to take part in the survey).

The results of Proposition 5.2 allow us to establish the first main result of this section. In Theorem 5.3 we state that it is possible to strictly increase the total level of contribution at the non-zero equilibrium by imposing a minimum precision level. In particular, such an increase is monotonic in η in the domain $[0, \eta^*]$.

Theorem 5.3. *Suppose that Assumptions 1 and 2 are satisfied. The total level of contribution $G^*(N, \eta)$ at equilibrium $\boldsymbol{\lambda}^*(N, \eta)$ of the modified game $\Gamma(N, \eta)$ is a non-decreasing function of $\eta \in [0, \eta^*]$ and, in particular, it is increasing in $[\lambda_n^*, \eta^*]$.*

It is important to recall that, without loss of generality, with Assumption 2 we considered an original game without free-riders. Under this assumption, in this subsection, we have shown, up to a certain value of the minimum contribution level, the existence of a NE still without free-riders in which the increase of the total contribution level is strictly positive and it is monotonic in the choice of the parameter η. In the more general setting, this translates into the existence of a NE where all the individuals who were previously contributing do not free-ride. However, the uniqueness of such a NE is not assured and the equilibrium prediction could suggest the rise of a suboptimal NE with potentially many (more) free-riders. In the following, we address this problem by refining the NE definition and by showing that, up to a more restrictive value of the minimum contribution level η, $\boldsymbol{\lambda}^*(N, \eta)$ is in fact the only equilibrium which is likely to arise by implementing the corresponding modified game.

5.3 The Nash equilibria refinement

By introducing a minimum precision level, we face the problem of possibly having a multiplicity of Nash equilibria. In Section 3, we have provided the definition of a Nash equilibria refinement, based on the potential nature of our game, that we called the PMNE. Such equilibria are the most likely to arise in an experimental implementation and they represent a robust way of predicting the outcome of such a strategic situation. For these reasons, in the following we investigate whether, for some choices of the minimum contribution level η, we can ensure the existence of a unique PMNE and whether this solution coincides or not with the non-zero component NE of the previous subsection. In Definition 5.4, we introduce a new parameter $\bar{\eta}$ smaller than or equal to η^* such that, as we see in Proposition 5.5, if we set a minimum precision level strictly smaller than $\bar{\eta}$, the unique NE s.t. there are no free-riders is also the unique PMNE and then it represents the only one that is likely to emerge through an implementation of the model.

Definition 5.4. *Suppose that Assumptions 1 and 2 are satisfied. We define the parameter $\bar{\eta}$ as the smallest parameter in $(\lambda_n^*, \eta^*]$, if there exists one, s.t.,*

$$\Phi|_{\lambda_n=0}(\boldsymbol{\lambda}^{M0}(N \setminus \{n\}, \bar{\eta})) = \Phi(\boldsymbol{\lambda}^M(N, \bar{\eta})), \tag{7}$$

where $\boldsymbol{\lambda}^{M0}(N \setminus \{n\}, \bar{\eta})$ is the unique global maximum of the potential function Φ defined on the restricted domain $[\bar{\eta}, 1]^{n-1} \cup \{0\}$ and $\boldsymbol{\lambda}^M(N, \bar{\eta})$ is the unique global maximum of Φ on $[\bar{\eta}, 1]^n$. If such a parameter does not exist, we impose $\bar{\eta} = \eta^*$.

To better understand this definition, we observe that the parameter $\bar{\eta}$ is defined as the smallest minimum contribution level at which the maximum value of the potential function is the same by letting individual n free to choose whether to contribute or not or by forcing him to contribute zero. Of course, for values of the minimum level of contribution in $[0, \lambda_n^*]$, the maximum value of the potential function is unique and this condition will never be satisfied.

Proposition 5.5. *Suppose that Assumptions 1 and 2 are satisfied. For any $\eta \in (0, \bar{\eta})$, $\boldsymbol{\lambda}^*(N, \eta) = \boldsymbol{\lambda}^M(N, \eta)$ is the unique PMNE of the modified game $\Gamma(N, \eta)$.*

Complementary to the results of Proposition 5.2 and Theorem 5.3, Proposition 5.5 states the second main result of this section. We have shown in Proposition 5.2 how, introducing a minimum precision level which is smaller than η^*, we ensure that an equilibrium such that no individual is free riding (in the more general case, such that no individual who would contribute in the original game is not free-riding in the modified game either) still exists, and such an equilibrium is unique, i.e., any other equilibrium is such that some individuals contribute zero. By further bounding the choice of the minimum level of contribution to a parameter smaller than $\bar{\eta}$, we ensure in Proposition 5.5 that such an equilibrium is the unique potential maximizer and, as such, the unique one which is likely to arise implementing the model. Finally, Theorem 5.3 ensures that, by imposing such a minimum precision level, the total level of contribution at the PMNE, is strictly higher, compared to the one of the original game. We conclude by summing up that choosing a parameter η as close as possible to $\bar{\eta}$ represents, up to our knowledge, the best choice to maximize the total contribution level. Moreover we conjecture, as it is still an open question, that any choice of the parameter bigger than η^* will not produce any equilibrium of the game in which no one who is contributing in the original game is free-riding or with a bigger total level of contribution.

In Corollary 5.6, we illustrate how there could exist other PMNE for choices of the parameter η which are too large, i.e., when trying to force the individuals to some contribution levels which are too high.

Corollary 5.6. *When $\eta = \eta^*$, for each $j \in N$ s.t. $p'_j(\cdot) = p'_n(\cdot)$ there exists a NE $\boldsymbol{\nu}$ of $\Gamma(N, \eta^*)$ s.t., $\nu_i = \lambda_i^*(N, \eta^*) = \lambda_i^*(N \setminus \{n\}, \eta^*)$ for each $i \neq j$ and $\nu_j = 0$. Such an equilibrium is such that $\Phi(\boldsymbol{\nu}) = \Phi(\boldsymbol{\lambda}^*(N, \eta^*))$.*

From Corollary 5.6 we may observe that when implementing the game $\Gamma(N, \eta^*)$, any strategy vector which corresponds to $\boldsymbol{\lambda}^*(N, \eta^*)$, but s.t. one of the individuals who have the largest contribution cost deviates to zero, is still a NE (in particular, it is still a PMNE if the original contribution vector is a PMNE). This is because, by Definition 5.1, η^* is the contribution level s.t. not only her individual utility, but also the potential function does not vary when individual n deviates. However, such a PMNE is inefficient in terms of total level of contribution, as it is s.t. $n-1$ individuals are contributing the same and 1 individual who was contributing strictly more than zero is now free-riding.

Some additional results on the public good economy with incentives under the assumption of homogeneous contribution costs are presented in Section 6.2. These results will be of particular interest for the example presented in the following Section 6.

6 Further Results for an Homogeneous Public Goods Economy

Given the game $\Gamma(N)$, now we assume that the individuals utility is still defined by (1), but it is s.t. $p_i(\cdot) = p(\cdot)$ for each $i \in N$. When the game $\Gamma(N)$ is homogeneous, we denote it by $\Gamma(n)$, i.e., as dependent on the number of individuals and not on their identity. We observe that in this special case, Assumption 1 naturally holds. The homogeneous case allows us to understand in more details the equilibrium structure of the original and of the modified models in the particular situation in which all the individuals have the same contribution cost. If this hypothesis may sound restrictive, it is also important to underline that this analysis is also of particular interest when implementing a public goods game in which the contribution costs of the individuals are not known precisely, and only an estimation or an average of the population is available.

6.1 The original homogeneous public goods economy

Under the homogeneity assumption, we present the first result, that simply follows as a corollary of Theorem 4.1.

Corollary 6.1. *Let the game $\Gamma(n)$ be homogeneous, the unique NE $\boldsymbol{\lambda}^*$ is s.t. $\lambda_i^* = \lambda^* > 0$ for each $i \in N$.*

This corollary states in particular that, under the homogeneity assumption, also Assumption 2 naturally holds, i.e., at equilibrium, all the individuals are contributing to the public good.

When the game $\Gamma(n)$ is homogeneous, we adopt the notation $\lambda^*(n)$ and $G^*(n)$ to denote the individual and the total equilibrium level of contribution respectively. This helps us emphasize that, under the hypothesis of homogeneity, the equilibrium level does not depend on the specific identity of the set of the individuals N, but only on its cardinality.

The second result, that we present in Proposition 6.2, could be partially shown as a corollary of Proposition 4.2. However, we present it in a more complete version, as the result of a comparative statics analysis of the individual contribution and the total level of contribution at equilibrium, studying how these quantities vary when changing the parameter n. In particular, we show that the individual contribution goes to zero when the number of individuals goes to infinity, and that the total level of contribution goes to infinity.

Proposition 6.2. *Let the game $\Gamma(n)$ be homogeneous. The individual contribution at equilibrium $\lambda^*(n)$ and the total level of contribution at equilibrium $G^*(n)$ satisfy:*

(i) *$\lambda^*(n)$ is a non-increasing function of the number n of individuals, and it is s.t. $\lim_{n\to+\infty} \lambda^*(n) = 0$;*

(ii) *$G^*(n) = n\lambda^*(n)$ is an increasing function of the number n of individuals, and it is s.t. $\lim_{n\to+\infty} G^*(n) = +\infty$.*

As a consequence of Proposition 6.2, we observe that, even if the total level of contribution goes to infinity when the number of individuals goes to infinity, it still goes slower than the optimal total level of contribution, as it holds that

$$\frac{n}{G^*(n)} = \frac{n}{n\lambda^*(n)} = \frac{1}{\lambda^*(n)} \longrightarrow +\infty.$$

To conclude the analysis of our original model in the homogeneous case, we also observe that the inequalities of Corollary 4.3 hold now strictly, i.e., at equilibrium, the individual utility and the social welfare are strictly bigger when a new individual enters the game.

6.2 The modified homogeneous public goods economy

Similarly to the previous section, we now detail the results of Propositions 5.2 and 5.5 and Theorem 5.3 for the modified homogeneous game, that we denote by $\Gamma(n, \eta)$. First, in Definitions 6.3 and 6.4, we introduce a parameter $\eta^*(n)$ and a parameter $\bar{\eta}(n)$, which are the analogous of parameters η^* and $\bar{\eta}$ defined in Definitions 5.1 and 5.4 for the heterogeneous case.

Definition 6.3. *We define the parameter $\eta^*(n)$ as the smallest parameter in $(\lambda^*(n), 1]$, if there exists one, s.t.*

$$h(n\eta^*(n)) - h((n-1)\eta^*(n)) = p(\eta^*(n)). \tag{8}$$

If such a parameter does not exist, we impose $\eta^*(n) = 1$.

Definition 6.4. *We define the parameter $\bar{\eta}(n)$ as the smallest parameter in $(\lambda^*(n), \eta^*(n)]$, if there exists one, s.t.*

$$h((n-1)\lambda^*(n-1)) - (n-1)p(\lambda^*(n-1)) = h(n\bar{\eta}(n)) - np(\bar{\eta}(n)). \tag{9}$$

If such a parameter does not exist, we impose $\bar{\eta}(n) = \eta^*(n)$.

In Theorem 6.5, we provide a more detailed analysis of the NE structure for the homogeneous case. In particular, in this subcase we are able to prove the conjecture such that, when the minimum precision level is too high (bigger than $\eta^*(n)$), at equilibrium some individuals necessarily free-ride.

Theorem 6.5. *Let the game $\Gamma(n)$ be homogeneous, and let $\boldsymbol{\lambda}^*(n)$ be the unique Nash equilibrium of the original game $\Gamma(n)$. It holds:*

(i) *if $n = 1$, then for any $\eta \in (0, \eta^*(n)]$, $\Gamma(n, \eta)$ has a unique Nash equilibrium $\lambda^*(n, \eta) = \max\{\lambda^*(1), \eta\}$. When $\eta = \eta^*(n)$, there are two NE in 0 and in $\eta^*(n)$, while for any $\eta \in (\eta^*(n), 1]$, there is a unique NE in 0;*

(ii) *if $n > 1$, then*

(iia) *for any* $\eta \in (0, \eta^*(n))$, $\Gamma(N, \eta)$ *has a unique NE* $\boldsymbol{\lambda}^*(n, \eta)$ *s.t.* $\lambda_i^*(n, \eta) > 0$ *for each* $i \in N$. *This equilibrium is s.t.,* $\lambda_i^*(n, \eta) = \lambda^*(n, \eta)$ *for each* $i \in N$, *with*

$$\lambda^*(n, \eta) = \begin{cases} \lambda^*(n) & \text{if } 0 \leq \eta \leq \lambda^*(n) \\ \eta & \text{if } \lambda^*(n) < \eta < \eta^*(n); \end{cases} \tag{10}$$

In particular, if $\eta \in (0, \bar{\eta}(n))$, *such an equilibrium is the unique PMNE.*

(iib) *for* $\eta = \eta^*(n)$, $\Gamma(n, \eta^*(n))$ *has at least* $n+1$ *NE: an equilibrium* $\boldsymbol{\lambda}^*(n, \eta^*(n))$ *s.t.* $\lambda_i^*(n, \eta^*(n)) = \eta^*(n)$ *for each* $i \in N$, *and* n *equilibria* $\boldsymbol{\nu}^1, \ldots, \boldsymbol{\nu}^n$ *s.t. for each* $j = 1 \ldots n$ $\nu_i^j = \eta^*(n)$ *for each* $i \neq j$ *and* $\nu_j^j = 0$. *In particular, if* $\boldsymbol{\lambda}^*(n, \eta^*(n))$ *is a PMNE, then all the* $\boldsymbol{\nu}^j$ *are PMNE for each* $j = 1 \ldots n$.

(iic) *for any* $\eta \in (\eta^*(n), 1]$, *there does not exist a Nash equilibrium* $\bar{\boldsymbol{\lambda}}$ *of* $\Gamma(n, \eta)$ *s.t.* $\bar{\lambda}_i > 0$ *for each* $i \in N$.
Moreover, assume that $\eta \leq \eta^*(1)$ *and let* m_η *be such that* $\eta \in (\eta^*(m_\eta + 1), \eta^*(m_\eta)]$. *Then, any profile such that* m_η *individuals contribute* $\max(\lambda^*(m_\eta), \eta)$ *and the rest gives zero is a NE; and there exists no NE with* s *individuals contributing positively for any* $s > m$. *Finally, if* $\eta > \eta^*(1)$, *the only equilibrium is* $(0, \cdots, 0)$.

The previous theorem (iic) shows that, for any value of $\eta \in (\eta^*(m_\eta + 1), \eta^*(m_\eta)]$, one can at best get m_η individuals contributing $\eta^*(m_\eta)$ at a NE. That is, the maximal total contribution at a NE is $m\eta^*(m)$ for some $m \leq n$. Hence, in all cases where $m\eta^*(m)$ is an increasing function of m, which is the case for instance in the data analytics example, one cannot do better in terms of total public good contribution by setting $\eta > \eta^*(n)$.

The last result of this section shows the monotonicity of the parameter $\eta^*(n)$, which is a non-increasing function of the parameter n, going to zero for an infinitely large number of individuals.

Corollary 6.6. *The function* $\eta^*(n)$ *is non-increasing in the number* n *of individuals and decreasing if* $\eta^*(n) < 1$, *and it is s.t.* $\lim_{n \to +\infty} \eta^*(n) = 0$.

It follows that also the parameter $\bar{\eta}(n)$ goes to zero when the number of individuals goes to infinity. As a consequence, similarly to what we observed for the homogeneous case without minimum contribution level, also the modified game performs worst than the maximal level of contribution n for an infinite number of individuals, as it holds that

$$\frac{n}{n\lambda^*(n, \bar{\eta}(n))} = \frac{n}{n\bar{\eta}(n)} = \frac{1}{\bar{\eta}(n)} \longrightarrow +\infty.$$

However, as we show for the particular case in Section 7, the improvement can still be very significant.

7 A Data Analytics Project as a Public Goods Economy

Our results provide a widely applicable method to increase the provision of a public good above voluntary contributions, simply by restricting the individuals' strategy spaces. This method is attractive by its simplicity and, consequently, its applicability. In this section, we illustrate the model and our results on our motivating example about privacy, where the economy is given by the data analytics research project we illustrated in Section 3.3.

We explicit the results and we run some simulations in the special case where the privacy cost is homogeneous, i.e., s.t. $p_i^E(\cdot) = p^E(\cdot)$ for each $i \in N$, and monomial and the estimation cost is linear; i.e., we assume that the cost function in (4) has the form

$$J_i(\lambda_i, \boldsymbol{\lambda}_{-i}) = c\lambda_i^k + \sigma_M^2(\boldsymbol{\lambda}), \tag{11}$$

where $c \in (0, \infty)$ and $k \geq 2$ are constants. This corresponds to a cost of contribution $p_i^E(\lambda_i) = c\lambda_i^k$, and a public good utility $h^E(\boldsymbol{\lambda}) = -\frac{1}{\sum_{i \in N} \lambda_i}$. Note that, without loss of generality, in the linear estimation cost, we omit the constant factor (adding a constant to the cost does not modify the game solutions) as well as the slope factor (adding it would give an equivalent game with constant c rescaled). According to the previous notation, we denote the resulting game by $\Gamma^E(n)$.

For this special case, we explicit the theoretical results we presented in the previous sections. In fact, we can determine the equilibrium precision $\lambda^*(n)$, the optimal minimum precision level $\eta^*(n)$ and the contribution level that minimizes the social cost $\lambda^e(n)$ in closed form. In the special case of costs given by (11), the equilibrium precision chosen by the agents in the game $\Gamma^E(n)$ simplifies to:

$$\lambda^*(n) = \begin{cases} \left(\dfrac{1}{ckn^2}\right)^{\frac{1}{k+1}} & \text{if } \left(\dfrac{1}{ckn^2}\right)^{\frac{1}{k+1}} \le 1/\sigma^2 \\ 1/\sigma^2 & \text{if } \left(\dfrac{1}{ckn^2}\right)^{\frac{1}{k+1}} > 1/\sigma^2, \end{cases} \tag{12}$$

the upper bound of the minimum precision level to guarantee the existence of a non-zero NE is given by:

$$\eta^*(n) = \begin{cases} \left(\dfrac{1}{cn(n-1)}\right)^{\frac{1}{k+1}} & \text{if } \left(\dfrac{1}{cn(n-1)}\right)^{\frac{1}{k+1}} \le 1/\sigma^2 \\ 1/\sigma^2 & \text{if } \left(\dfrac{1}{cn(n-1)}\right)^{\frac{1}{k+1}} > 1/\sigma^2, \end{cases}$$

if $n \ge 2$, and it is equal to $1/\sigma^2$ otherwise, while the upper bound to guarantee that such an equilibrium is the unique PMNE is given by $\bar{\eta}(n) = \eta^*(n)$. The contribution level that minimizes the social cost is:

$$\lambda^e(n) = \begin{cases} \left(\dfrac{1}{ckn}\right)^{\frac{1}{k+1}} & \text{if } \left(\dfrac{1}{ckn}\right)^{\frac{1}{k+1}} \le 1/\sigma^2 \\ 1/\sigma^2 & \text{if } \left(\dfrac{1}{ckn}\right)^{\frac{1}{k+1}} > 1/\sigma^2. \end{cases} \tag{13}$$

We may conclude that the optimal choice for the data analyst in order to maximize the total level of contribution (which is equivalent to minimizing the variance of the estimation) and to have every individual participating, is to set up a minimum contribution level as close as possible to $\eta^*(n)$. How these quantities vary as functions of n, together with the corresponding variance of the estimation, is presented in Figure 1.

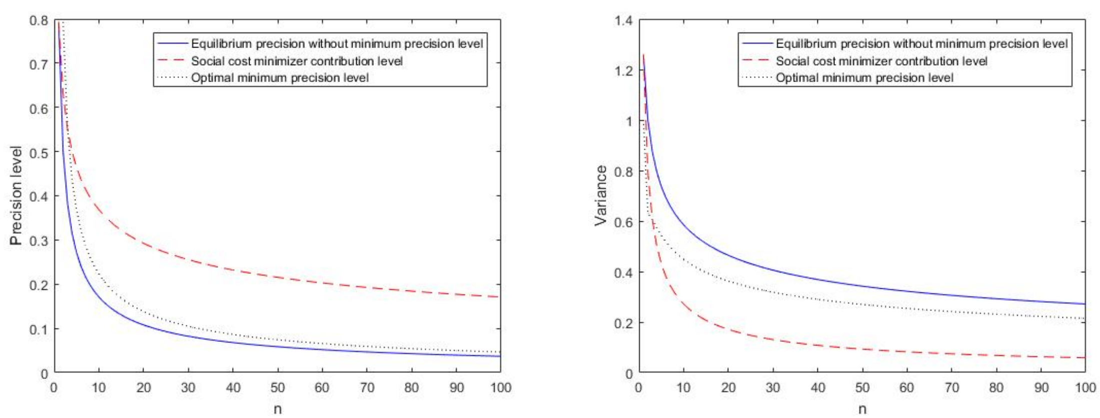

Figure 1: Precision levels and corresponding variance of the estimation when varying $n = 1, \dots, 100$, for $k = 2$ and $c = 1$.

Writing explicitly the previous quantities, we can more easily analyze the properties of our public goods game and its modification that we have already discussed in the previous sections for the general case. In particular, we observe that when c increases, i.e., when the individuals are more concerned about privacy, they choose at equilibrium a smaller precision level $\lambda^*(n)$. Further, the minimum precision level $\eta^*(n)$ becomes smaller if the individuals are more sensitive about the protection of their data. Finally, we have $\lambda^*(n) < \eta^*(n)$ for each $n \in \mathbb{N}^*$, and both of these quantities decrease and go to zero when n increases and goes to $+\infty$.

Interestingly, the closed-form expressions that we have for this special case allow us to analyze the rate of decrease of the variance (equivalent to the rate of increase of the contribution), and to quantify the improvement that can be achieved by imposing a minimum precision level. For n large enough (such that both $\lambda^*(n)$ and $\eta^*(n)$ are strictly smaller than $1/\sigma^2$), the variance at equilibrium level $\lambda^*(n)$ of game Γ^E is given by

$$\sigma_M^2(\boldsymbol{\lambda}^*(n)) = \frac{1}{n\left(\frac{1}{ckn^2}\right)^{\frac{1}{k+1}}},$$

while by setting up a minimum precision level η as close as possible to $\eta^*(n)$, we can reach a variance given by

$$\sigma_M^2(\boldsymbol{\lambda}^*(n,\eta)) = \frac{1}{n\left(\frac{1}{cn(n-1)}\right)^{\frac{1}{k+1}}} + \epsilon,$$

with ϵ arbitrarily small.

Both appear to have the same rate of decrease in $n^{\frac{-k+1}{k+1}}$ which is smaller than n^{-1} but becomes closer to n^{-1} as k tends to infinity. Intuitively, as the privacy cost becomes closer to a step function, the equilibrium precision level becomes less dependent on the number of agents so that we get closer to the case of averaging iid random variables of fixed variance. Consequently, for n large enough, the improvement is given by a factor:

$$\frac{\sigma_M^2(\boldsymbol{\lambda}^*(n))}{\sigma_M^2(\boldsymbol{\lambda}^*(n,\eta))} = \left(\frac{kn}{n-1}\right)^{\frac{1}{k+1}} > 1, \tag{14}$$

which asymptotically becomes constant:

$$\frac{\sigma_M^2(\boldsymbol{\lambda}^*(n))}{\sigma_M^2(\boldsymbol{\lambda}^*(n,\eta^*(n)))} \sim_{n\to\infty} k^{\frac{1}{k+1}}. \tag{15}$$

Interestingly, we notice that this ratio of variances (characterizing the improvement when setting the optimal minimum precision level) depends on k, but not on c. (This holds even before the asymptotic regime, as long as n is large enough such that both $\lambda^*(n)$ and $\eta^*(n)$ are strictly smaller than $1/\sigma^2$.)

Figure 2 illustrates the asymptotic improvement ratio (15) for different values of k. We observe that it is bounded, it goes to 1 for large k's and it is in the range of $25-30\%$ improvement for values of k around $2-10$. Given that the ratio (14) converges towards its asymptote from above, this asymptotic improvement represents a lower bound of the improvement the analyst can achieve by implementing our mechanism with any finite number n of agents.

As already discussed, the goal of minimizing the variance is often not at odds with the maximization of the social welfare, i.e., in our example, with the minimization of the social cost $W^E(\boldsymbol{\lambda}) = \sum_{i\in N} J_i^E(\lambda)$. We observe that having a minimum precision level that minimizes both the variance and the social cost is an exception that may be true only in the case in which $\lambda^*(n) = \lambda^e(n) = 1/\sigma^2$. In all the other cases, our modification of the structure of the game will lead to a second best equilibrium. In Figure 3 we represent the social cost function, and the corresponding precision levels. We observe that indeed, imposing a minimum precision level does not yield a minimum of the social cost, but it still improves (i.e., decreases) the social cost compared to the game without incentive.

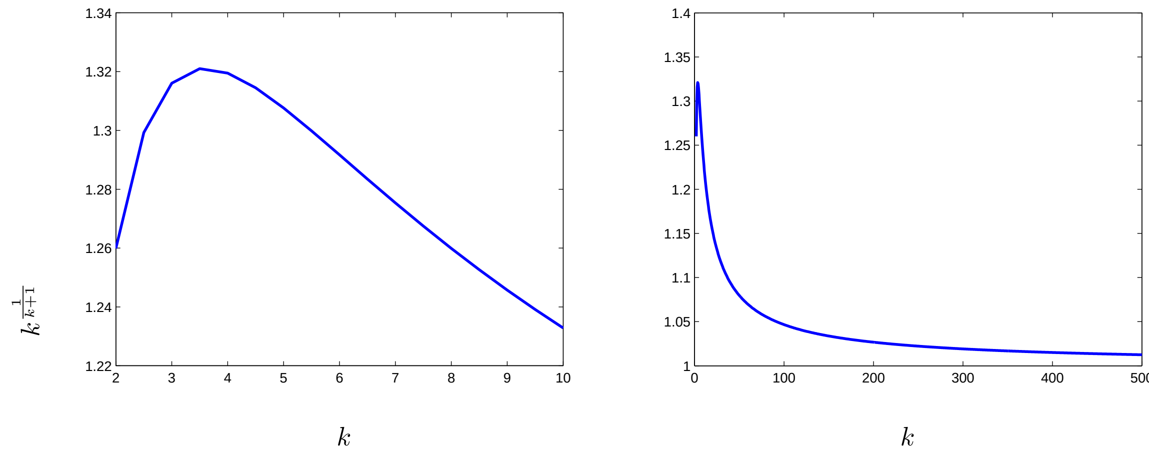

Figure 2: Asymptotic improvement of the estimation choosing the optimum precision level $\eta^*(n)$ for values of $k = 2, \ldots, 10$ and for values of $k = 2, \ldots, 500$.

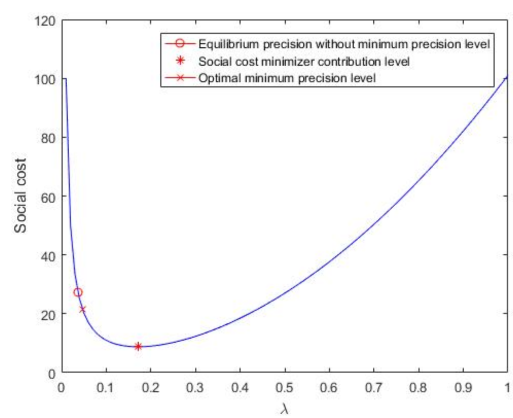

Figure 3: Social cost function for $n = 100$, $k = 2$ and $c = 1$.

8 Concluding Remarks

In this paper, we propose a mechanism to enhance voluntary contribution in a public goods economy, which consists in restricting the individuals' strategy set by imposing a minimum contribution level while still allowing free-riding (i.e., choosing a zero contribution level). We show that this intervention can strictly increases the total level of contributions at a unique potential maximizer Nash equilibrium, while it does not incentivize any (new) free-riding. Protocols that impose some form of a minimum contribution level are already widely applied in areas such as data analytics (where a minimum amount of information is often required to participate) or charitable giving (in which voluntary benefactors are usually asked to contribute a minimum amount such as the price of a charity raffle ticket). Such protocols, however, were not investigated—which is our focus. Our work formalizes them and explains theoretically their efficacy. It also identifies the minimum contribution level as an interesting mechanism design tool, in particular for cases where other types of intervention (e.g., monetary) are not possible. Our theoretical results also provide an operational way to choose the minimum contribution level optimally. Interestingly, they also predict that setting it too high may result in free-riding and in a decrease of the total contribution level—which could be used to explain situations where excessive free-riding occurs.

Our work is motivated by applications in which the public good is unambiguously good—which is standard in economics. Even in our main application of data analytics, our motivation originated from cases such as medical research where the result of the data analysis is indeed clearly a public *good*. Our proposed mechanism may, however, also be relevant in situations where data collection is more controversial such as the massive data collection of online platforms. Indeed, the consequences of the collection of billions of individuals' data through online applications and integrated technologies such as the Internet of Things have become a major concern in the last decades. Despite years of alternated scandals (e.g., the *Cambridge Analytica*[11] scandal) and regulations (e.g., the European *General Data Protection Regulation*), the debate on how to regulate and optimize data markets for both users and data analysts is still active (see, e.g., some recent papers by Gan [2019] and by Bergemann et al. [2022]). Such markets are particularly controversial because of the often opposite interests of the different agents (users and data analysts), because of the unclear position of some of those agents (e.g., users often have a contradictory attitude known as *Privacy Paradox* [Liao et al., 2021]), and because of the externalities a user sharing her data has on other users [Acemoglu et al., 2022]. In this context, our mechanism could be relevant to increase the contribution of users who benefit from the data collection, although the theoretical model would require modifications to fit the aforementioned market's characteristics. Recall, however, that our mechanism lets users free-ride if they choose to, which would continue to happen for highly privacy sensitive users. Finally, we reemphasize that our primary motivation was not this type of data collection, but rather data analytics problems with clear benefits to society such as medical research.

A natural follow up of this paper is to validate and complement the theoretical results through a behavioral study. Most of the existing mechanisms and approaches for improving the performances of public goods economies have been tested experimentally in an increasingly vast literature on behavioral economics and sociology of public goods provision, see e.g., [Bohm, 1972, Holt and Laury, 1997, Lugovskyy et al., 2017, Marwell and Ames, 1979, Zelmer, 2003]. A laboratory experiment could investigate two key points: first validate the theoretical prediction of our paper, and second test whether it is possible to push even farther the theoretical threshold represented by the minimum contribution level. Indeed, the reputation effect, often present in models of voluntary contribution to a public good, can be boosted by the higher contribution of other cooperators. In particular, we may expect that the presence of a minimum contribution threshold would be perceived by individuals as a sort of psychological default option. The experimental literature has already shown that subjects are reluctant to switch from a default option [Liu and Riyanto, 2017], and how this effect may boost cooperation. Finally, another interesting direction of future analysis is to investigate theoretically whether our model is stable to small income redistribution among contributors and how the crowding-out generated by a government grant would get affected by the minimum contribution requirement.

[11] Cambridge Analytica inferred information about more than 50 million Facebook users, which it deployed for designing personalized political messages and advertising in the Brexit referendum and 2016 US presidential election.

A Proofs

A.1 Proof of Theorem 4.1

Game $\Gamma(N)$ is a potential game with a concave potential function Φ (defined as in (2)) and where the strategy sets are closed intervals of the real line. It follows that a strategy profile is a Nash equilibrium if and only if it maximizes the potential function. As the potential function Φ is strictly concave on the convex set on which it is defined, it has a unique global maximum $\boldsymbol{\lambda}^* \in [0,1]^n$ which coincides with the unique Nash equilibrium of $\Gamma(N)$ and such an equilibrium is strict. In particular, the equilibrium $\boldsymbol{\lambda}^*$ is such that for each $i \in N$, λ_i^* satisfies the following KKT conditions

$$\begin{cases} h'(G^*(N)) - p_i'(\lambda_i^*) + \psi_i^* - \phi_i^* = 0 \\ \psi_i^* \lambda_i^* = 0 \quad \phi_i^*(\lambda_i^* - 1) = 0, \quad \psi_i^*, \phi_i^* \geq 0, \end{cases} \tag{16}$$

where by $G^*(N)$ we denote the total contribution level at equilibrium. We may observe by (16) that, owing to the assumption that $h'(0) > p_i'(0)$ for at least one $i \in N$, this equilibrium is s.t., $\boldsymbol{\lambda}^* \neq (0, \dots, 0)$.

As the term $h'(G^*(N))$ is the same for each $i \in N$, it immediately follows that, if the p_i's are s.t. $p_1'(\lambda) \leq ... \leq p_n'(\lambda)$, for each $\lambda \in [0,1]$, then $\lambda_n^* \leq \ldots \leq \lambda_1^*$.

A.2 Proof of Proposition 4.2

(i) We suppose by contradiction that there exists $i \in N$ s.t. $0 \leq \lambda_i^*(N) < \lambda_i^*(N \cup \{n+1\}) \leq 1$. Because of the strict convexity of p_i,

$$p_i'(\lambda_i^*(N)) < p_i'(\lambda_i^*(N \cup \{n+1\})),$$

and, from the KKT conditions in (16), and as $\lambda_i^*(N) < 1$ and $\lambda_i^*(N \cup \{n+1\}) > 0$,

$$h'(G^*(N)) \leq p_i'(\lambda_i^*(N)) < p_i'(\lambda_i^*(N \cup \{n+1\})) \leq h'(G^*(N \cup \{n+1\})).$$

From the first and from the last term, and because of the concavity of h, it follows that

$$G^*(N) > G^*(N \cup \{n+1\}). \tag{17}$$

If the total contribution without individual $n+1$ is strictly bigger than after her entrance, this implies that there exists at least one individual $j \in N$ which verifies the opposite inequality compared to i, i.e., such that $1 \geq \lambda_j^*(N) > \lambda_j^*(N \cup \{n+1\}) \geq 0$. Following the same reasoning than before, we conclude that $G^*(N) < G^*(N \cup \{n+1\})$ and this contradicts (17).

(ii) From (i), we know that $\lambda_i^*(N \cup \{n+1\}) \leq \lambda_i^*(N)$ for each $i \in N$. In particular, if the equality holds for each $i \in N$, the thesis follows trivially, as $G^*(N \cup \{n+1\}) = G^*(N) + \lambda_{n+1}^*(N \cup \{n+1\})$, with $\lambda_{n+1}^*(N \cup \{n+1\}) \geq 0$. On the contrary, if there exists $i \in N$ s.t. $\lambda_i^*(N \cup \{n+1\}) < \lambda_i^*(N)$, then we can conclude that the inequality holds strictly, following the same reasoning than in (i).

A.3 Proof of Corollary 4.3

It is sufficient to observe that, at equilibrium,

$$\begin{aligned} U_i^*(N \cup \{n+1\}) &= h(G^*(N \cup \{n+1\})) - p_i(\lambda_i^*(N \cup \{n+1\})) \\ &\geq h(G^*(N)) - p_i(\lambda_i^*(N)) \\ &= U_i^*(N) \end{aligned}$$

where the inequality follows because of Proposition 4.2.

A.4 Proof of Proposition 5.2

We observe that the modified game $\Gamma(N,\eta)$ is still a potential game, with potential function Φ as in (2), but defined on the restricted domain $\big[\{0\}\cup[\eta,1]\big]^n$. Differently from $\Gamma(N)$, the strategy sets of the modified game are not closed intervals of the real line and the equilibria may not coincide with the global maxima of the potential function.

(i) When $|N|=1$, the potential function and the utility function coincide. Then, it still holds that a strategy profile is a Nash equilibrium if and only if it is a global maximum of Φ. If $\eta\le\lambda^*$, the unique global maximum of Φ is still λ^*. If $\eta\in(\lambda^*,\eta^*)$, the unique global maximum is η. If $\eta=\eta^*$, there are two global maxima in 0 and in η^*. If $\eta\in[\eta^*,1]$, then there is again a unique maximum of Φ but in 0.

(ii) Now, let $|N|>1$. For a better reading of the following of the proof, we present it organized in 6 steps.
Step 1 First,we observe that, as the unique NE of the original game $\Gamma(N)$ $\boldsymbol{\lambda}^*$ is strict, in particular it holds that for each $i\in N$,

$$h(G^*(N))-p_i(\lambda_i^*)>h(G^*(N)-\lambda_i^*),$$

i.e., by deviating to zero any individual strictly decreases her utility.

Step 2 Given a subset of the individuals $S\subseteq N$, with $s=|S|$, and for each $\eta\in[0,1]$, we define a new modified game $\Gamma'(S,\eta)=\langle S,[\eta,1]^s,(U_i)_{i\in S}\rangle$, where the utility function U_i is defined by (1) for each individual $i\in S$, but it is now restricted on the domain $[\eta,1]$. $\Gamma'(S,\eta)$ is still a potential game, with potential function Φ as in (2), defined on the restricted domain $\{0\}^{n-s}\cup[\eta,1]^s$[12]. As for the original game $\Gamma(N)$ and the modified game $\Gamma(N,\eta)$, we observe that also the game $\Gamma'(S,\eta)$ is a potential game with a concave potential function and where, similarly to $\Gamma(N)$ but differently from $\Gamma(N,\eta)$, the strategy sets are closed intervals of the real line. It follows that a strategy profile is a Nash equilibrium if and only if it maximizes the potential function. Moreover, the potential function is strictly concave on the convex set on which it is defined, then it has a unique global maximum $\boldsymbol{\lambda}^M(S,\eta)\in[\eta,1]^s$, which coincides with the unique Nash equilibrium of $\Gamma'(S,\eta)$. In particular, for each $\eta\in(0,1]$, $\boldsymbol{\lambda}^M(N,\eta)$ is the unique Nash equilibrium of $\Gamma'(N,\eta)$.

Step 3 We define η^* as in (5). We show that $\boldsymbol{\lambda}^M(N,\eta^*)$ is a Nash equilibrium of $\Gamma(N,\eta^*)$. At first, observe that no individual has incentives to deviate to a quantity in $[\eta^*,1]$, because $\boldsymbol{\lambda}^M(N,\eta^*)$ is a Nash equilibrium of the game $\Gamma'(N,\eta^*)$. It remains to be shown that no individual has incentives to deviate to zero. Individual n does not have incentives by (5) or by (6). For any other individual $i\ne n$, s.t. $\lambda_i^M(N,\eta^*)=\eta^*$, if agent n, who is the most privacy concerned, does not have incentives to deviate from η^*, that is still valid for i. For any other agent $i\ne n$, s.t. $\lambda_i^M(N,\eta^*)>\eta^*$, as i does not have incentives to deviate to η^*, then, because of the concavity of the utility function, she cannot have incentives to deviate to 0.

Step 4 We observe that for each $\eta\in[0,\eta^*)$, $\boldsymbol{\lambda}^M(N,\eta)$ is a Nash equilibrium of $\Gamma(N,\eta)$. This follows trivially from Theorem 4.1 when $\eta\in(0,\lambda_n^*]$. Moreover, when $\eta\in(\lambda_n^*,\eta^*)$, we can repeat the same reasoning of Step 3, with the only difference that now individual n (and any other individual contributing η) is always strictly better off by contributing rather than free-riding, and then the inequality is always strict.

Step 5 We observe that for any $\eta\in(0,\eta^*]$, $\boldsymbol{\lambda}^M(N,\eta)$ is the unique Nash equilibrium of $\Gamma(N,\eta)$ s.t. each individual has a non-zero contribution, i.e., s.t. $\lambda_i^M(N,\eta)>0$ for each $i\in N$. To show that, it is sufficient to observe that an equilibrium of $\Gamma(N,\eta)$ s.t. each individual has a non-zero contribution, is also an equilibrium of the game $\Gamma'(N,\eta)$ (as if an individual does not have incentives to deviate in $\{0\}\cup[\eta,1]$, she does not have incentives to deviate in the restricted strategy set $[\eta,1]$ either), and the equilibrium of $\Gamma'(N,\eta)$ is unique.

Step 6 We show that any other NE $\bar{\boldsymbol{\lambda}}$ of $\Gamma(N,\eta)$ is s.t. $\bar\lambda_i=\lambda_i^M(S,\eta)$ for each $i\in S$, where $S=\{i\in N|\bar\lambda_i\ne 0\}$. Let $\eta\in(\lambda_n^*,\eta^*]$ and let $\bar{\boldsymbol{\lambda}}=\bar{\boldsymbol{\lambda}}(N)$ be a Nash equilibrium of $\Gamma(N,\eta)$ s.t.

[12]For simplicity, we make an abuse of notation by assuming that the zero components correspond to the first $n-s$ individuals.

$\bar{\lambda}_i(N) = 0$ for at least one $i \in N$. We define S as the set of individuals who contribute non-zero at this equilibrium. As, by definition of NE, none of the individuals has incentives to deviate, and, in particular, none of the individuals in S has incentives to deviate in $[\eta, 1]$, it follows that $\bar{\boldsymbol{\lambda}}(S)$, i.e., the vector of the non-zero elements of the vector $\bar{\boldsymbol{\lambda}}(N)$, is a Nash equilibrium of the game $\Gamma'(S, \eta)$ defined in step 2 of this proof. As we have seen, this equilibrium is unique and it coincides with the unique Nash equilibrium $\boldsymbol{\lambda}^M(S, \eta)$ of the game $\Gamma(S, \eta)$ s.t. no individual is free-riding. Trivially, we can conclude that $\boldsymbol{\lambda}^*(N, \eta)$ is the unique NE of $\Gamma(N, \eta)$ s.t. $\lambda_i^*(N, \eta) > 0$ for all $i \in N$.

A.5 Proof of Theorem 5.3

Remember that we assumed that the equilibrium contribution $\boldsymbol{\lambda}^*(N)$ of $\Gamma(N)$ is such that $\lambda_n^*(N) < 1$, i.e., it is such that the total level of contribution is not optimal.

When $\eta \in [0, \lambda_n^*(N)]$, $\boldsymbol{\lambda}^*(N, \eta) = \boldsymbol{\lambda}^*(N)$ and, consequently, G^* is constant, i.e., $G^*(N, \eta) = G^*(N)$.

When $\eta \in (\lambda_n^*(N), \eta^*]$, individual n is contributing at equilibrium $\lambda_n^*(N, \eta) = \eta > \lambda_n^*(N)$. We show now that for $\eta^* \geq \eta_2 > \eta_1 \geq \lambda_n^*$, $G^*(N, \eta_2) > G^*(N, \eta_1)$. Assume, by contradiction, that $G^*(N, \eta_2) \leq G^*(N, \eta_1)$. It follows that

$$h'(G^*(N, \eta_2)) \geq h'(G^*(N, \eta_1)), \tag{18}$$

because of the concavity of the public good utility function h. Moreover, as the total level of contribution with η_2 is smaller than or equal to the level of contribution with η_1, but we know that individual n has a strictly larger level of contribution, it follows that there exists $i \in N$ who contributes strictly less, i.e., s.t.,

$$\eta_2 \leq \lambda_i^*(N, \eta_2) < \lambda_i^*(N, \eta_1) \leq 1. \tag{19}$$

By equation (19), because of the convexity of the cost of contribution function p_i and from the KKT conditions for the potential function of the modified game $\Gamma^*(N, \eta)$, it follows that

$$h'(G^*(N, \eta_2)) \leq p_i'(\lambda_i^*(N, \eta_2)) < p_i'(\lambda_i^*(N, \eta_1)) \leq h'(G^*(N, \eta_1))$$

and this contradicts equation (18).

A.6 Proof of Proposition 5.5

Let $\bar{\eta}$ be defined as in Definition 5.4, $\eta \in (\lambda_n^*, \bar{\eta})$, and $\boldsymbol{\lambda}^*(N, \eta) = \boldsymbol{\lambda}^M(N, \eta)$ be the unique non-zero components Nash equilibrium of the modified game $\Gamma(N, \eta)$, which is the unique maximizer of Φ on $[\eta, 1]^n$. As $\eta < \bar{\eta}$, it follows that

$$\Phi(\boldsymbol{\lambda}^M(N, \eta)) > \Phi|_{\lambda_n = 0}(\boldsymbol{\lambda}^{M0}(N \setminus \{n\}), \eta). \tag{20}$$

As $\boldsymbol{\lambda}^{M0}(N \setminus \{n\})$ is the potential maximizer of function Φ restricted on the domain $[\bar{\eta}, 1]^{n-1} \cup \{0\}$, it also holds that

$$\Phi|_{\lambda_n = 0}(\boldsymbol{\lambda}^{M0}(N \setminus \{n\}), \eta) \geq \Phi|_{\lambda_i = 0, \forall i \in N \setminus S}(\boldsymbol{\lambda}^{M0}(S, \eta)). \tag{21}$$

From (20) and (21), it follows that $\boldsymbol{\lambda}^*(N, \eta) = \boldsymbol{\lambda}^M(N, \eta)$ is the unique PMNE of the modified game $\Gamma(N, \eta)$.

A.7 Proof of Corollary 5.6

First, we observe that it holds

$$h(G^*(N, \eta^*)) - p_j(\lambda_j^*(N, \eta^*)) \geq h(G^*(N, \eta^*) - \lambda_j^*(N, \eta^*)) \quad \forall j \in N \tag{22}$$

because, as we proved in Proposition 5.2, $\boldsymbol{\lambda}^*(N, \eta^*)$ is a NE of $\Gamma(N, \eta^*)$, and then no individual has incentives to deviate to zero. Now, we suppose by contradiction that the vector $\boldsymbol{\nu}$ s.t., $\nu_i = \lambda_i^*(N, \eta^*) = \lambda_i^*(N \setminus \{n\}, \eta^*)$

for each $i \neq n$ and $\nu_n = 0$ is not a NE of $\Gamma(N, \eta^*)$. This means that there exists an individual $j \in N \setminus \{n\}$ for whom it is convenient to deviate, i.e., s.t., it holds that

$$h(G^*(N,\eta^*) - \eta^*) - p_j(\lambda_j^*(N,\eta^*)) < h(G^*(N,\eta^*) - \eta^* - \lambda_j^*(N,\eta^*)). \tag{23}$$

From Equations (22) and (23), it follows that

$$\begin{aligned} & h(G^*(N,\eta^*)) - h(G^*(N,\eta^*) - \lambda_j^*(N,\eta^*)) \geq p_j(\lambda_j^*(N,\eta^*)) \\ & > h(G^*(N,\eta^*) - \eta^*) - h(G^*(N,\eta^*) - \eta^* - \lambda_j^*(N,\eta^*)) \end{aligned}$$

and this cannot hold because of the concavity of h. It follows that $\boldsymbol{\nu}$ is a NE of $\Gamma(N, \eta^*)$.

Second, we observe that $\boldsymbol{\nu}$ provides the same value of the potential function than $\boldsymbol{\lambda}^*(N, \eta^*)$. Then, if the second one is a potential maximizer, the first one is as well.

A.8 Proof of Corollary 6.1

When the game $\Gamma(N)$ is homogeneous, the potential function Φ in (2) is a symmetric function on a symmetric domain. As a consequence, the unique maximum is also symmetric, i.e., $\lambda_i^* = \lambda^*$ for each $i \in N$. In particular, as the equilibrium vector is a non zero vector, we have that $\lambda^* > 0$.

A.9 Proof of Proposition 6.2

(i) Firstly, from Equation (16), remembering we are now in a homogeneous case, we observe that $\lambda^* > 0$ is the unique solution of the following fixed point problem

$$\lambda = g(n, \lambda), \tag{24}$$

where function $g : \mathbb{N}_+ \times (0, 1] \to [0, +\infty]$ is defined for each $\lambda \in (0, 1]$ and for each $n \in \mathbb{N}_+$ as

$$g(n,\lambda) = \min\left\{(p')^{-1}(h'(n\lambda)), 1\right\}. \tag{25}$$

We consider the problem with the parameter n defined on the real interval $[1, +\infty]$. For each $n \in [1, +\infty]$, g is continuous in λ. Moreover, function g is monotonic non-increasing in n. Indeed,

$$\frac{\partial g}{\partial n} = \frac{\lambda}{p''((p')^{-1}(h'(n\lambda)))} h''(n\lambda) < 0$$

for each λ internal solution, and g is identically 1 otherwise. Applying Corollary 1 of Milgrom and Roberts [1994], the unique fixed point $\lambda^*(n) > 0$ is non-increasing in n and, consequently, $\lim_{n\to+\infty} \lambda^*(n) \geq 0$ is well defined.

Secondly, we observe that, for each $\lambda > 0$,

$$\lim_{n\to+\infty} g(n,\lambda) = 0$$

pointwise. Indeed, $\lim_{x\to+\infty} h'(x) = 0$, $p'(0) \geq 0$ and p' is strictly monotonic.

If we suppose by contradiction that $\lim_{n\to+\infty} \lambda^*(n) = a > 0$, then, from Equation (24), it follows that

$$0 < \lim_{n\to+\infty} \lambda^*(n) = \lim_{n\to+\infty} g(n, \lambda^*(n)) = \lim_{n\to+\infty} g(n, a) = 0.$$

(ii) By Proposition 4.2-(ii), we know that $G^*(n+1) \geq G^*(n)$ for each $n \in N$ and, in particular, that the inequality holds strictly as at equilibrium in the homogeneous case any new individuals is not a free-rider. We suppose by contradiction that $\lim_{n\to+\infty} G^*(n) < +\infty$. It follows that $\lim_{n\to+\infty} h'(n\lambda^*(n)) > 0$ and then, by (25), that the solution to the fixed point problem is s.t. $\lim_{n\to+\infty} \lambda^*(n) > 0$, and this contradicts the second statement of Proposition 6.2-(i).

A.10 Proof of Theorem 6.5

(i) The trivial case when $n = 1$ is unchanged from the analogous case of Theorem 5.2-(i).

(ii) Now, let $n > 1$. For each $\eta \in (0, 1]$, let $\boldsymbol{\lambda}^M(n, \eta)$ be the unique global maximum of Φ on $[\eta, 1]^n$. When the individuals are homogeneous, the potential function Φ is a symmetric function that we are maximizing on a symmetric domain. As a consequence, the maximum is s.t., $\lambda_i^M(n, \eta) = \lambda^M(n, \eta)$ for each $i \in N$, with

$$\lambda^M(n,\eta) = \begin{cases} \lambda^*(n) & \text{if } 0 < \eta \le \lambda^*(n) \\ \eta & \text{if } \lambda^*(n) < \eta \le 1. \end{cases} \tag{26}$$

(iia) As a particular case of Proposition 5.2-(ii), for each $\eta \in (0, \eta^*(n)]$, where $\eta^*(n)$ is defined as in Definition 6.3, $\boldsymbol{\lambda}^*(n, \eta) = \boldsymbol{\lambda}^M(n, \eta)$ is the unique NE of the modified game $\Gamma(n, \eta)$ s.t. no individual is free riding. As a particular case of Proposition 5.5, for each $\eta \in (0, \bar{\eta}(n)]$, where $\bar{\eta}(n)$ is defined as in Definition 6.4, $\boldsymbol{\lambda}^*(n, \eta)$ is the unique PMNE of the modified game $\Gamma(n, \eta)$.

(iib) This result follows as the homogeneous case of the statement of Corollary 5.6.

(iic) For each $\eta \in (\eta^*(n), 1]$, we show that there does not exist a NE $\bar{\boldsymbol{\lambda}}$ of $\Gamma(n, \eta)$ s.t. $\bar{\lambda}_i > 0$ for each $i \in N$. We assume $\eta^*(n) < 1$. First we observe that, with a reasoning similar to the one in the proof of Proposition 5.2, A.4-Step 5, whenever we have a non-zero components NE of $\Gamma(n, \eta)$, this has to be a NE of the corresponding game Γ', an then a maximum of the potential function Φ restricted on the domain $[\eta, 1]^n$. It follows that, because of the symmetry, the only possible candidate non-zero components NE is the vector $\bar{\boldsymbol{\lambda}} = \boldsymbol{\eta} = (\eta, \ldots, \eta)$. Second, we observe that by definition, $\eta^*(n)$ is the smallest solution of the following fixed point problem

$$p(\eta) = h(n\eta) - h((n-1)\eta) \tag{27}$$

or, equivalently, of

$$\frac{p(\eta)}{\eta} = \frac{h(n\eta) - h((n-1)\eta)}{\eta} \tag{28}$$

where n is a fixed parameter. Then, we show that $\eta^*(n)$ is, in fact, the unique solution of (28). Indeed, because of the concavity of h, we have that

$$\frac{h(y) - h(x)}{y - x}$$

is a decreasing function both in y and in x and then, in particular, the right term of (28) is decreasing in η. Moreover, because of the convexity of p, we have that

$$\frac{p(y) - p(x)}{y - x}$$

is an increasing function both in y and in x and then, in particular, the left term of (28) is increasing in η (when $y = \eta$, $x = 0$ and $p(0)$ is a constant). It follows that the fixed point problem in (28) has at most one solution.

In particular, applying Corollary 1 of Milgrom and Roberts [1994], the unique (i.e. smallest) fixed point $\eta^*(n) > 0$ is non-increasing in n. Moreover, if we assume that $\eta^*(n) = \eta^*(n-1) = \eta < 1$, we obtain from (27) that

$$h(n\eta) - h((n-1)\eta) = h((n-1)\eta) - h((n-2)\eta),$$

which contradicts the fact that h is strictly concave. Hence we conclude that $\eta^*(n)$ is strictly decreasing in n.

Third, we know by Theorem 4.1 that $\boldsymbol{\lambda}^*(n)$ is a SNE of $\Gamma(n)$, i.e., that $p(\lambda^*(n)) < h(n\lambda^*(n)) - h((n-1)\lambda^*(n))$. Moreover, we have just observed that $\eta^*(n)$ verifies the equality by definition whenever we have that $\eta^*(n) < 1$. As the solution of the previous fixed point problem, if it exists, is unique, it follows that

$$p(\eta) > h(n\eta) - h((n-1)\eta) \tag{29}$$

for each $\eta \in (\eta^*(n), 1]$. Equation (29) translates the fact that, when the individuals adopt the strategy $\boldsymbol{\eta} = (\eta, \ldots, \eta)$, with $\eta \in (\eta^*(n), 1]$, each individual in N is strictly better of by deviating to zero, and then the only candidate non-zero components NE cannot be a NE.

Now assume that $\eta \leq \eta^*(1)$ and let $m_\eta \in \{1, \cdots, n-1\}$ be the integer such that $\eta \in (\eta^*(m_\eta + 1), \eta^*(m_\eta)]$. Such an integer exists and is unique due to the fact that $\eta^*(n)$ is strictly decreasing in n.

Note that by definition of m_η, we see that it is possible that $\eta^*(m_\eta) = 1$ (in which case $\eta^*(m) = 1$ for all $m \leq m_\eta$ but those values of m never correspond to any $m_{\eta'}$ for any η'); but we necessarily have $\eta^*(m_\eta + 1) < 1$.

Assume that there exists a NE $\bar{\boldsymbol{\lambda}}$ of $\Gamma(N, \eta)$ where s individuals in set S contribute positively, i.e., $\bar{\lambda}_i > 0$ for all $i \in S$ and $\bar{\lambda}_i^* = 0$ for all $i \in N \setminus S$, where $s = |S|$. First note that, following arguments from the proof of Proposition 5.2, $(\bar{\lambda}_i)_{i \in S}$ is equal to $\boldsymbol{\lambda}^M(S, \eta)$ which is symmetric. We denote by $\bar{\lambda}$ the value of $\bar{\lambda}_i$ for $i \in S$ and we have $\bar{\lambda} = \lambda^*(s)$ is $\eta < \lambda^*(s)$ and $\bar{\lambda} = \eta$ otherwise.

Next, we observe that we must have $s \leq m_\eta$. Indeed, assume that $s > m_\eta$. Then $\lambda^*(s) \leq \eta^*(s) \leq \eta^*(m_\eta + 1) < \eta$, so that $\bar{\lambda} = \eta$. In that case, an individual $i \in S$ switching from $\bar{\lambda}$ to zero contribution would gain $p(\bar{\lambda}) - \big(h(s\bar{\lambda}) - h((s-1)\bar{\lambda})\big)$, which is strictly positive since, using arguments similar to the ones introduced above, we can show:

$$\begin{aligned}
\frac{p(\bar{\lambda})}{\bar{\lambda}} = \frac{p(\eta)}{\eta} &> \frac{p(\eta^*(m+1))}{\eta^*(m+1)} \\
&= \frac{h((m+1)\eta^*(m+1)) - h(m\eta^*(m+1))}{\eta^*(m+1)} \\
&\geq \frac{h(s\eta^*(m+1)) - h((s-1)\eta^*(m+1))}{\eta^*(m+1)} \\
&\geq \frac{h(s\bar{\lambda}) - h((s-1)\bar{\lambda})}{\bar{\lambda}}.
\end{aligned}$$

This contradicts the fact that $\bar{\boldsymbol{\lambda}}$ is a NE.

We now show that for $s = m_\eta$, the profile $\bar{\boldsymbol{\lambda}}$ with $\bar{\lambda}_i = \max(\lambda^*(s), \eta) \in (\eta^*(m_\eta + 1), \eta^*(m_\eta)]$ for all $i \in S$ is indeed a NE. First note that, by definition of $\eta^*(m_\eta + 1)$, the individuals outside S (i.e., giving zero) have no incentive to deviate. Indeed, if $i \in N \setminus S$ deviates to $\lambda' \geq \eta > \eta^*(m_\eta + 1)$, his gain is

$$\begin{aligned}
h((s+1)\lambda') - h(s\lambda') - p(\lambda') &= \lambda' \frac{h((s+1)\lambda') - h(s\lambda')}{\lambda'} - \lambda' \frac{p(\lambda')}{\lambda'} \\
&\leq \lambda' \frac{h((s+1)\eta^*(m_\eta+1)) - h(s\eta^*(m_\eta+1))}{\eta^*(m_\eta+1)} - \lambda' \frac{p(\eta^*(m_\eta+1))}{\eta^*(m_\eta+1)} \\
&= 0,
\end{aligned}$$

where we used again the convexity of p and concavity of h as derived above. A similar reasoning shows that individuals in S have no incentive to deviate to zero, hence $\bar{\boldsymbol{\lambda}}$ is a NE.

A.11 Proof of Corollary 6.6

The proof on the first part of this corollary is already included in the proof of Theorem 6.5. It remains to observe that, as a consequence of the monotonicity, $\lim_{n \to +\infty} \eta^*(n) \geq 0$ is well defined. Indeed,

$\lim_{x \to +\infty} h(nx) - h((n-1)x) = 0$, $p(0) \geq 0$ and p is strictly monotonic. If we suppose by contradiction that $\lim_{n \to +\infty} \eta^*(n) = a > 0$, then, from Equation (27), it follows that

$$\begin{aligned}
0 &< p(a) \\
&= \lim_{n \to +\infty} p(\eta^*(n)) \\
&= \lim_{n \to +\infty} h(n\eta^*(n)) - h((n-1)\eta^*(n)) \\
&= \lim_{n \to +\infty} h(na) - h((n-1)a) \\
&= 0.
\end{aligned}$$

References

J. M. Abowd and I. M. Schmutte. An economic analysis of privacy protection and statistical accuracy as social choices. *American Economic Review*, 109:171–202, 2019.

D. Acemoglu, A. Makhdoumi, A. Malekian, and A. Ozdaglar. Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics*, 14:218–256, 2022.

A. Acquisti and C. Fong. An experiment in hiring discrimination via online social networks. *Management Science*, 66(3):1005–1024, 2020.

A. Acquisti and J. Grossklags. Privacy and rationality in individual decision making. *IEEE Security & Privacy*, 3:26–33, 2005.

A. R. Admati and M. Perry. Joint projects without commitment. *The Review of Economic Studies*, 58: 259–276, 1991.

A. C. Aitken. On least squares and linear combinations of observations. In *Proceedings of the Royal Society of Edinburgh*, 1935.

G. A. Akerlof. A theory of social custom, of which unemployment may be one consequence. *Quarterly Journal of Economics*, 94:749–775, 1980.

I. Altman. *The environment and social behavior: privacy, personal space, territory, and crowding.* ERIC, 1975.

J. Andreoni and T. C. Bergstrom. Do government subsidies increase the private supply of public goods? *Public Choice*, 88:295–308, 1996.

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

D. Bergemann, A. Bonatti, and T. Gan. The economics of social data. *RAND Journal of Economics*, 53: 263–296, 2022.

T. Bergstrom, L. Blume, and H. R. Varian. On the private provision of public goods. *Journal of Public Economics*, 29:25–49, 1986.

P. Bohm. Estimating demand for public goods: An experiment. *European Economic Review*, 3:111–130, 1972.

Y. Bramoullé and R. Kranton. Public goods in networks. *Journal of Economic Theory*, 135:478–494, 2007.

R. M. Burlando and F. Guala. Heterogeneous agents in public goods experiments. *Experimental Economics*, 8:35–54, 2005.

C. B. Cadsby and E. Maynes. Voluntary provision of threshold public goods with continuous contributions: experimental evidence. *Journal of Public Economics*, 71:53–73, 1999.

M. Chessa, J. Grossklags, and P. Loiseau. A short paper on incentives to share private information for population estimates. In *Proceedings of the 19th International Conference on Financial Cryptography and Data Security (FC)*, 2015a.

M. Chessa, J. Grossklags, and P. Loiseau. A game-theoretic study on non-monetary incentives in data analytics projects with privacy implications. In *Proceedings of the 2015 IEEE 28th Computer Security Foundations Symposium (CSF)*, pages 90–104, 2015b.

E. H. Clarke. Multipart pricing of public goods. *Public Choice*, 11:17–33, 1971.

R. Cornes and T. Sandler. Easy riders, joint production, and public goods. *Economic Journal*, pages 580–598, 1984.

P. Courtois and G. Haeringer. Environmental cooperation: ratifying second-best agreements. *Public Choice*, 151:565–584, 2012.

L. Damschroder, J. Pritts, M. Neblo, R. Kalarickal, J. Creswell, and R. Hayward. Patients, privacy and trust: Patients' willingness to allow researchers to access their medical records. *Social Science & Medicine*, 64:223–235, 2007.

S. Duca, D. Helbing, and H. H. Nax. Assortative matching with inequality in voluntary contribution games. *Computational Economics*, 52:1029–1043, 2018.

J. C. Duchi, M. I. Jordan, and M. J. Wainwright. Minimax optimal procedures for locally private estimation. *Journal of the American Statistical Association*, 113(521):182–201, 2018.

E. Fehr and S. Gächter. Cooperation and punishment in public goods experiments. *The American Economic Review*, 90:980–994, 2000.

T. Gan. Optimal contracts for data generators, 2019. Working paper, available at SSRN: `https://ssrn.com/abstract=3477080` (Last revised Jan 2022).

N. Gast, S. Ioannidis, P. Loiseau, and B. Roussillon. Linear regression from strategic data sources. *ACM Transactions on Economics and Computation*, 8, 2020.

T. Groves and J. Ledyard. Optimal allocation of public goods: A solution to the "free rider" problem. *Econometrica*, 45:783–809, 1977.

A. Guido, A. Robbett, and R. Romaniuc. Group formation and cooperation in social dilemmas: A survey and meta-analytic evidence. *Journal of Economic Behavior and Organization*, 159:192–209, 2019.

B. Herrmann, C. Thöni, and S. Gächter. Antisocial punishment across societies. *Science*, 319:1362–1367, 2008.

C. A. Holt and S. K. Laury. Classroom games: Voluntary provision of a public good. *Journal of Economic Perspectives*, 11:209–215, 1997.

B. Huberman, E. Adar, and L. Fine. Valuating privacy. *IEEE Security & Privacy*, 3:22–25, 2005.

L. Hurwicz. On informationally decentralized systems. In R. Radner and C. McGuire, editors, *Decision and Organization*, pages 297–336. North Holland, Amsterdam, 1972.

S. Ioannidis and P. Loiseau. Linear regression as a non-cooperative game. In *Proceedings of the 9th International Conference on Web and Internet Economics (WINE)*, pages 277–290, 2013.

A. Kajii and S. Morris. The robustness of equilibria to incomplete information. *Econometrica*, 65:1283–1309, 1997.

N. Kass, M. Natowicz, S. Hull, R. Faden, L. Plantinga, L. Gostin, and J. Slutsman. The use of medical records in research: What do patients want? *Journal of Law, Medicine & Ethics*, 31:429–433, 2003.

P. Kollock. Social dilemmas: The anatomy of cooperation. *Review of Sociology*, 24:183–214, 1998.

J. O. Ledyard. Public goods: A survey of experimental research. In Kagel and Roth, editors, *Handbook of Experimental Economics*. Princeton University Press, Princeton, 1995.

J. O. Ledyard and T. R. Palfrey. Voting and lottery drafts as efficient public goods mechanisms. *Review of Economic Studies*, 61:327–355, 1994.

J. O. Ledyard and T. R. Palfrey. The approximation of efficient public good mechanisms by simple voting schemes. *Journal of Public Economics*, 83:153–171, 2002.

G. Liao, Y. Su, J. Ziani, A. Wierman, and J. Huang. The privacy paradox and optimal bias-variance trade-offs in data acquisition. In *Proceedings of the 22nd ACM Conference on Economics and Computation (EC)*, page 689, 2021. Full paper available as arXiv:2105.14262.

J. Liu and Y.E. Riyanto. The limit to behavioral inertia and the power of default in voluntary contribution games. *Social Choice and Welfare*, 48:815–835, 2017.

L. Lugovskyy, D. Puzzello, A. Sorensen, J. Walker, and A. Williams. An experimental study of finitely and infinitely repeated linear public goods games. *Games and Economic Behavior*, 102:286–302, 2017.

N. K. Malhotra, S. S. Kim, and J. Agarwal. Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. *Information Systems Research*, 15:336–355, 2004.

G. Marwell and R. E. Ames. Experiments on the provision of public goods i: Resources, interests, group size, and the free rider problem. *American Journal of Sociology*, 84:1335–1360, 1979.

M. A. McCain. Information as property and as a public good: Perspectives from the economic theory of property rights. *The Library Quarterly: Information, Community, Policy*, 58:265–282, 1988.

M. McGinty and G. Milam. Public goods provision by asymmetric agents: experimental evidence. *Social Choice and Welfare*, 40:1159–1177, 2013.

J. Mikians, L. Gyarmati, V. Erramilli, and N. Laoutaris. Crowd-assisted search for price discrimination in e-commerce: First results. In *Proceedings of the Conference on Emerging Networking Experiments and Technologies (CoNEXT)*, pages 1–6, 2013.

P. Milgrom and J. Roberts. Comparing equilibria. *American Economic Review*, 84:441–459, 1994.

M. Milinski, D. Semmann, H-J Krambeck, and J. Marotzke. Stabilizing the earth's climate is not a losing game: supporting evidence from public goods experiments. *Proceedings of the National Academy of Sciences of the United States of America*, 103 11:3994–3998, 2006.

D. Monderer and L.S. Shapley. Potential games. *Games and Economic Behavior*, 14:124–143, 1996.

H. H. Nax, R. O. Murphy, and D. Helbing. Stability and welfare of 'merit-based' group-matching mechanisms in voluntary contribution game, 2014. Available at SSRN: http://dx.doi.org/10.2139/ssrn.2404280.

N. Nikiforakis. Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, 92:91–112, 2008.

A. Oliveira, R. Croson, and C. Eckel. One bad apple? heterogeneity and information in public good provision. *Experimental Economics*, 18:116–135, 2015.

D. Oyama and O. Tercieux. Iterated potential and robustness of equilibria. *Journal of Economic Theory*, 144:1726–1769, 2009.

T. R. Palfrey and H. Rosenthal. Participation and the provision of discrete public goods: a strategic analysis. *Journal of Public Economics*, 24(2):171–193, 1984.

A. Ramachandran and A. Chaintreau. The network effect of privacy choices. In *Proceedings of NetEcon*, pages 59–62, 2015.

D. G. Rand and M. A. Nowak. Human cooperation. *Trends in cognitive sciences*, 17:413–425, 2013.

R. Roberts. A positive model of private charity and wealth transfers. *Journal of Political Economy*, 92: 136–148, 1984.

M. Robling, K. Hood, H. Houston, R. Pill, J. Fay, and H. Evans. Public attitudes towards the use of primary care patient record data in medical research without consent: A qualitative study. *Journal of Medical Ethics*, 30:104–109, 2004.

B. Roussillon, N. Gast, P. Loiseau, and P. Mertikopoulos. Asymptotic degradation of linear regression estimates with strategic data sources. In *Proceedings of the 33rd International Conference on Algorithmic Learning Theory (ALT)*, 2022.

P. A. Samuelson. The pure theory of public expenditure. *The Review of Economics and Statistics*, 36: 387–389, 1954.

K. Sigmund, C. Hauert, and M. Nowak. Reward and punishment. *Proceedings of the National Academy of Sciences of the United States of America*, 98:10757–10762, 2001.

S. Spiekermann, J. Grossklags, and B. Berendt. E-privacy in 2nd generation e-commerce: Privacy preferences versus actual behavior. In *Proceedings of the 3rd ACM Conference on Electronic Commerce*, pages 38–47, 2001.

R. Sugden. On the economics of philanthropy. *Economic Journal*, pages 341–350, 1982.

A. van de Kragt, J. Orbell, and R. Dawes. The minimal contributing set as a solution to public goods problems. *American Political Science Review*, 77:112–122, 1983.

J. Van Huyck, R. Battalio, , and R. Beil. Tactic coordination games, strategic uncertainty, and coordination failure. *American Economic Review*, 80:234–248, 1990.

H. R. Varian. Sequential contributions to public goods. *Journal of Public Economics*, 53:165–186, 1994.

H. R. Varian. Beyond big data. *Business Economics*, 49:27–31, 2014.

S. Warren and L. Brandeis. The Right to Privacy. *Harvard Law Review*, pages 193–220, 1890.

A. Westin. *Privacy and freedom.* Atheneum, New York, 1970.

J. Wieringa, P.K. Kannan, X. Ma, T. Reutterer, H. Risselada, and B. Skiera. Data analytics in a privacy-concerned world. *Journal of Business Research*, 122:915–925, 2021.

D. T. Young. Voluntary purchase of public goods. *Public Choice*, 38:73–85, 1982.

J. Zelmer. Linear public goods experiments: A meta-analysis. *Experimental Economics*, 6:299–310, 2003.

Acknowledgements

The authors would like to thank Jens Grossklags, Heinrich Nax, Galina Schwartz and Paolo Zeppini for helpful comments and suggestions on earlier drafts of this paper. This work has been partially supported by the Multidisciplinary Institute in Artificial Intelligence MIAI @ Grenoble Alpes (ANR-19-P3IA-0003), by the French National Research Agency (ANR) under Grant ANR-20-CE23-0007, and by the Alexander von Humboldt Foundation.

DOCUMENTS DE TRAVAIL GREDEG PARUS EN 2017
GREDEG Working Papers Released in 2017

2017-01 Lauren Larrouy & Guilhem Lecouteux
Mindreading and Endogenous Beliefs in Games

2017-02 Cilem Selin Hazir, Flora Bellone & Cyrielle Gaglio
Local Product Space and Firm Level Churning in Exported Products

2017-03 Nicolas Brisset
What Do We Learn from Market Design?

2017-04 Lise Arena, Nathalie Oriol & Iryna Veryzhenko
Exploring Stock Markets Crashes as Socio-Technical Failures

2017-05 Iryna Veryzhenko, Etienne Harb, Waël Louhichi & Nathalie Oriol
The Impact of the French Financial Transaction Tax on HFT Activities and Market Quality

2017-06 Frédéric Marty
La régulation du secteur des jeux entre Charybde et Scylla

2017-07 Alexandru Monahov & Thomas Jobert
Case Study of the Moldovan Bank Fraud: Is Early Intervention the Best Central Bank Strategy to Avoid Financial Crises?

2017-08 Nobuyuki Hanaki, Eizo Akiyama, Yukihiko Funaki & Ryuichiro Ishikawa
Diversity in Cognitive Ability Enlarges Mispricing in Experimental Asset Markets

2017-09 Thierry Blayac & Patrice Bougette
Should I Go by Bus? The Liberalization of the Long-Distance Bus Industry in France

2017-10 Aymeric Lardon
On the Coalitional Stability of Monopoly Power in Differentiated Bertrand and Cournot Oligopolies

2017-11 Ismaël Rafaï & Mira Toumi
Pay Attention or Be Paid for Attention? Impact of Incentives on Allocation of Attention

2017-12 Gérard Mondello
Un modèle d'accident unilatéral: incertitude non-radicale et estimations différenciées

2017-13 Gérard Mondello
Lenders and Risky Activities: Strict Liability or Negligence Rule?

2017-14 Frédéric Marty
Économie des algorithmes et ordre concurrentiel : réflexions sur les abus d'exploitation et les collusions fondés sur des algorithmes de prix

2017-15 Agnès Festré, Pierre Garrouste, Ankinée Kirakozian & Mira Toumi
The Pen is Might Be Mightier than the Sword: How Third-party Advice or Sanction Impacts on Pro-environmental Behavior

2017-16 Edward Lorenz & Sophie Pommet
Innovation, Credit Constraints and National Banking Systems: A Comparison of Developing Nations

2017-17 Patrice Bougette & Christophe Charlier
Antidumping and Feed-In Tariffs as Good Buddies? Modeling the EU-China Solar Panel Dispute

2017-18 Nobuyuki Hanaki, Eizo Akiyama & Ryuichiro Ishikawa
Behavioral Uncertainty and the Dynamics of Traders' Confidence in their Price Forecasts

2017-19 Adnane Kendel, Nathalie Lazaric & Kevin Maréchal
What Do People 'Learn By Looking' at Direct Feedback on their Energy Consumption? Results of a Field Study in Southern France

2017-20 Nicolas Vallois & Dorian Jullien
Estimating Rationality in Economics: A History of Statistical Methods in Experimental Economics

2017-21 Nicolas Vallois & Dorian Jullien
Replication in Experimental Economics: A Historical and Quantitative Approach Focused on Public Good Game Experiments

2017-22 Benjamin Montmartin, Marcos Herrera & Nadine Massard
R&D Policy regimes in France: New Evidence from a spatio-temporal Analysis

2017-23 Thomas Boyer-Kassem & Sébastien Duchêne
On Discrimination in Health Insurance

2017-24 Michela Chessa & Patrick Loiseau
Enhancing Voluntary Contribution in a Public Goods Economy via a Minimum Individual Contribution Level