Public voluntary agreements as correlated equilibria of a subscription game : on the impact of a background regulatory threat

Anne-Sarah Chiambretto [1,*]

Aix-Marseille University (Aix-Marseille School of Economics), CNRS, & EHESS

Abstract

We develop a N-player subscription game, modified so as to represent the firms incentives to participate to an environmental public voluntary agreement (VA). Specifically, the latter is assumed to be preemptive, i.e. to occur under the threat of a mandatory regulation. We suggest the use of a correlation device to strengthen firms' participation decisions, formalized by the concept of correlated equilibrium (CE). It is first shown that any symmetrical mixed Nash equilibrium (NE) of the VA without a correlation device can be implemented by a regulator using the correlation device. Furthermore, we find that such a device not only solves the problem raised by multiplicity of NE, but also ensures that a higher expected aggregate payoff is reached for any given level of threat, t. Finally, we study the impact of the threat stringency on the set of CE. Our general results are illustrated in a specified example of pollution abatement model.

Keywords: collective voluntary agreements, pollution control, diffuse pollution, adoption costs, distributional effects, public goods, government policy.
JEL classification: C79, H23, H41, Q58.

[*]Corresponding author

Email address: anne-sarah@orange.fr (Anne-Sarah Chiambretto)

[1]P-mail address: GREQAM, Centre de la Vieille Charité, 2, rue de la Charité, 13236 Marseille, France; Tel.: +33 (0) 4 91 14 07 21; Fax: +33 (0) 4 91 90 02 27

Preprint submitted to DT-AMSE *May 2, 2016*

1. Introduction

. Collective voluntary approaches refer to commitments of groups of firms, possibly an entire industry, to cutting voluntarily their polluting emissions. When environmentally proactive behavior are motivated by a background regulatory threat, they fall under the particular denomination of voluntary-threat or preemptive voluntary approaches. In the growing theoretical literature on voluntary approaches to environmental protection, regulatory threats are only one among many reasons for firms to behave proactively. A second explanation is that it allows firms to build a green reputation and reach the new markets for eco-friendly goods and services (e.g. Arora and Gangopadhyay 1 and Cavaliere 5, or see Lyon and Maxwell 9 for a comprehensive review). Another, points out strategical effects related to market structure as a motive for overcompliance to environmental standard (e.g. Denicolò 8 and OECD 12 (sec. 5.4, 80)). But voluntary approaches do not only differ in terms of driving forces. Strongly depending on preexisting regulatory structures, they come in as many forms as institutional and legal backgrounds. A widely adopted classification distinguishes between three main categories[2], set on the basis of the regulatory agency involvement level (OECD 11) : (i) *public voluntary programs/agreements* (the agency elaborates engagements, to which the firms may voluntarily subscript), (ii) *negotiated agreements* (the engagements are collaboratively elaborated by the voluntary firms and the agency), and (iii) *unilateral commitments* (the engagements are elaborated by the voluntary firms solely). While, in theory, preemptive voluntary policies may possibly assume each of the three institutional forms aforementioned, in the field, they turn out to be most often related to policies scenarios of type (i) and (ii)[3], with collective liability rules.

The present work deals with industry-wide public voluntary agreements (VA), and focuses on firms' participation decision and its properties when the program features a correlation device. The game is reduced to a N-players subscription game such as defined in the standard literature on the private provision of a discrete public good (e.g. in Palfrey and Rosenthal 13). The public good to be provided is the non rival and non excludable regulatory gains that the firms derive from the mandatory regulation preemption arising from a succeeding public VA. These gains depend on the regulatory threat t, and the cost of providing the public good (i.e. the cost of satisfying the public VA policy requirement). Full participation is socially optimal as it is assumed the total cost of providing the public good decreases in the number of participating firms. Then, we use Aumann 2's correlated equilibrium concept to study the game outcome, with two possible interpretations for our results. If agents' strategy sets correspond to the sets of options naturally present in the game without a public VA policy, and agents coordinate on an exogenous random signal, then our results can be seen as a description of the role of inter-firm communication in self-regulation initiatives. If conversely, agents' strategy sets corresponds to a menu

[2]Some classifications may also include *direct negotiations*, or Cosean bargaining, as a fourth category, as well as information disclosure and eco-certification schemes as distinct kind of voluntary approaches.

[3]By construction, unilateral commitments are generally reputation-oriented (Borkey et al. 3)

defined by the public VA policy and the correlation device is implemented by a trustworthy regulator, then our results give rise to normative implications. We choose to develop the second interpretation. We find that, if instead offering participation to all the firms from the sector to be regulated, a regulator makes private recommendation to each firm, he can implement the best participation equilibrium that could have been possibly achieved in a case without private recommandation.

. Our results relate to the seminal Palfrey and Rosenthal 13 that studies, inter alia, the mixed Nash equilibria of a subscription game. Our findings differ from his in two points. While Palfrey and Rosenthal focus on the relationship between the set of Nash equilibria of subscription games, and the set of Nash equilibria of discrete private provision games *without* a refund, we use the concept of correlated equilibrium and focus on a slightly different game design. In the present work, the 'greed' motivation for free-riding can indeed be manipulated by the regulator through the tax threat, while assumptions on provision costs involve that the participation of a subgroup of players is socially inefficient. This specification reflects the participation incentives typically arising from preemptive public VAs with implicit collective liability. Otherwise, Cavaliere 6 studies the provision of discrete public good by correlated equilibria. Nevertheless, he restricts his study to the same general game *without* a refund as studied in Palfrey and Rosenthal 13. He finds that any convex combinations of pure strategies Nash equilibria are efficient correlated equilibria. He then introduces payoff externalities by assuming that both the consumption benefits and the contribution amount increases with the number of contributor, which still substantially differs from the game analyzed in the present work.

. This work is organized as follows. In section 2, we present the public VA, in a first place without a correlation device (also referred to as the non-mediated public VA), and second, with a correlation device (or mediated public VA). Section 3 characterizes, as benchmark results, the socially optimal allocation of the game, as well as the Nash equilibria of the non-mediated public VA. In section 4 we study the correlated equilibria of the mediated game, and show that, under a symmetry assumption, the unique mixed strategy Nash equilibrium can be implemented by a correlation device. Section 5 identifies the optimal correlated equilibrium of the game and provides a full study of the impact of the tax threat on collective welfare associated with the public VA. Finally, we numerically illustrate our results in a specified example of pollution abatement in section 6.

2. The model

2.1. The basic public VA

Consider a regulator willing to achieve a global regulation cap that involves the production[4] activities of N identical firms, indexed by $i = \{1, ..., N\}$. Such a cap may be to

[4]Our modified subscription game may as well apply to a more general context of externalities regulation, and the firms be simply called agents or players.

conform 60 % of a sector output with an efficiency standard (e.g. the ACEA agreement), or to reduce targeted agents' releases of a toxic chemical to some limit value (e.g. EPA's 33/50 Program), etc. To do so, the regulator offers each firm to participate to the target implementation by subscripting to a public VA. Firm i can accept ($s_i = 1$) or reject ($s_i = 0$) the offer, knowing that the participation modalities depend on the total number of participating firms N_p. Specifically, the regulator may consider the achievement of a given target by too few agents is cost-inefficient or not feasible. For instance, regarding the first example cited above, a participation rate of at least 0.6 is necessary for the cap to be implemented. Thus, if a participation threshold is not achieved, i.e. formally if

$$\sum_i^N s_i \equiv N_p < w,$$

where w is a positive integer, the VA fails and a collective tax t is enforced[5]. By collective, we mean the tax applies to each of the N agents whatever their individual willingness to undertake voluntary action may be. Conversely, if the participation threshold is achieved, the VA succeeds and the N_p participating firms implement the global target, equally sharing the implementation costs. Let us furthermore assume that, for any fixed cap level, the total cost of implementing the cap is a function $C : N_p \to \mathbb{R}$ decreasing in the number of participating firms:

$$c_i(N_p) = \begin{cases} \frac{C(N_p)}{N_p} & \text{if } N_p > 0 \\ 0 & \text{if } N_p = 0 \end{cases} \qquad \text{with } \frac{\partial C(N_p)}{\partial N_p} < 0, \tag{1}$$

where $c_i : N_p \to \mathbb{R}$ is the individual and symmetrical cost of implementing the cap for a participating firm i. This hypothesis may be seen as representing synergistic effects[6] in the target implementation and/or concavity of individual underlying objective functions, when rewritten as

$$\frac{\partial}{\partial N_p}\left(N_p c_i\left(N_p\right)\right) = N_p \frac{\partial c_i\left(N_p\right)}{\partial N_p} + c_i\left(N_p\right) < 0. \tag{2}$$

As for this second interpretation, think of agents as symmetrical firms with a concave profit function. Then, any cap set by the regulator that results in an output reduction does imply increasing individual average reduction costs[7]. Finally, let us define agent i's

[5]Dealing with symmetrical agents, t can be considered as the symmetrical individual tax payment resulting from the application of a given tax rate to the production or consumption activity to be regulated.

[6]See (ref.) for a literature review on synergistic effects in R&D.

[7]Let E denote the cap level and $c_i(E)$ the total implementation cost of E by agent i. If $c_i(E)$ is increasing in the cap stringency (or decreasing in E) and convex in E, we have that $\frac{c_i(E)}{E}$ is increasing in the cap stringency (or decreasing in E) and then $\frac{c_i(E/N_p)}{E/N_p}$ is decreasing in m. For a fixed cap level, we can rewrite $\frac{c_i(E/N_p)}{E/N_p} \equiv c_i\left(N_p\right) N_p$.

payoffs by

$$u_i(s) = \begin{cases} -c_i(N_p) & \text{if } s_i = 1,\ N_p \geq w \\ 0 & \text{if } s_i = 0,\ N_p \geq w \\ -t & \text{if } N_p < w \end{cases} \tag{3}$$

with $s \equiv (s_i)_i$ the strategy profile. We denote $c \equiv c_i(w)$, and assume that $t > c$.

2.2. The correlation device

Now suppose that instead of offering voluntary participation to each agent, the VA allows the regulator to privately recommend either $s_i = 0$ or $s_i = 1$, depending on i. Assume the recommendation is the result of a random selection process on the set of strategy profiles

$$S = \underset{i \in N}{\times} S_i, \quad s \in S,$$

with $S_i = \{1, 0\}$ the individual strategy sets, and $p \in \triangle(S)$ an underlying probability distribution where, by definition,

$$\triangle(S) = \left\{ (p_s)_{s \in S} \mid\ p_s \in [0,1], \sum_{s \in S} p_s = 1 \right\}.$$

While each agent is only told his individual strategy s_i in the randomly selected profile s, the probability distribution is assumed to be public knowledge. Suppose for example, that it is prealably announced by a trustworthy regulator. Again, each agent can freely follow or reject the recommendation, and the voluntary agreement succeeds as long as at least w agents participate. After a given distribution $p \in \triangle(S)$ is announced, agent i's *expected* utility is given by

$$\mathcal{U}_i(p) = \sum_{s \in S} p_s u_i(s). \tag{4}$$

This coordination device is a straightforward application of correlated strategies such as developed in Aumann 2.

Finally, as already stressed in the introduction regarding the voluntary public good provision literature, let us remark that theoretical interpretation of our framework are twofold. Here we choose to develop a VA model so the role of the regulator is well defined, and we can derive pragmatical implication for voluntary-threat policies design. Nevertheless, if strategy sets are said to be agents' natural action possibilities (as opposed to a menu defined by the voluntary program), and agents to correlate on an exogenous random signal, then the results should be reinterpreted as giving positive insights about self-regulation[8] and interfirms communication.

[8] i.e. "unilateral initiatives under which private parties, individuals or firms, voluntarily take actions to reduce pollution or protect natural resources without government involvement" Segerson 14.

3. Some preliminary results

In this section, we define the deterministic social optimum and study the set of NE of the basic game. In particular, we characterize the symmetric NE in mixed strategies (MNE), which will be used as a benchmark in next sections' analysis of the mediated public VA.

3.1. The deterministic optimum

Let us denote u_i^P (resp. u_i^{NP}) the utility function of participating (resp. non-participating) firms in the basic game. The regulator seeks to maximize the aggregate utility, which can be written as a maximisation program in the number of participating firms

$$\underset{N_p \in [0,N]}{max} U(N_p) := N_p u_i^P (N_p) + (N - N_p) u_i^{NP} (N_p), \tag{5}$$

where the objective simplifies to:

$$U(N_p) = \begin{cases} -Nt & \text{if } N_p < w \\ -N_p c_i(N_p) & \text{if } N_p \geq w. \end{cases} \tag{6}$$

First, remind that $t > c$, implying that all $N_p < w$ can not be solutions of the program. Then, it follows from the synergy hypotesis $C'(N_p) < 0$ that the solution is $N_p = N$, which corresponds to the full participation strategy profile $(1)_i$.

Let us remark that, since $U(N_p)$ is increasing in N_p, any participation profiles in S can be ranked in term of social welfare, according to the number of participating firms.

3.2. The multiplicity of NE

. As a preliminary result, we focus on the NE of the game. Let us first compute individual best responses $\Phi_i : S_{-i} \to \mathcal{P}(S_i)$ in the basic game, with $\mathcal{P}(S_i)$ the power set of S_i and

$$S_{-i} \equiv \times_{j \neq i} S_j, \quad s_{-i} \in S_{-i},$$

the set of feasible opponents' pure strategy profiles. From the payoff structure of the model and the definition of best response correspondences,

$$\underset{s_i \in \{0,1\}}{argmax} \quad u_i(s_{-i}, s_i),$$

we see that for all i:

$$\Phi_i (s_{-i}) = \begin{cases} \{0,1\} & \text{if } \sum_{j \neq i} s_j < w - 1 \\ 1 & \text{if } \sum_{j \neq i} s_j = w - 1 \\ 0 & \text{if } \sum_{j \neq i} s_j > w - 1 \end{cases}. \tag{7}$$

Remark that Φ_i can be redefined as a correspondence from $\mathbb{N}$ to $\mathcal{P}(S_i)$ since only the number of participating agents matters in the payoffs definition, and let us denote S^* the set of pure Nash equilibria (PNE), namely strategy profiles s^* defined by $\left(\Phi_i(s^*_{-i})\right)_i = s^*$. We then observe that any strategy profile such that $N_p < w - 1$ is a PNE, since a unilateral deviation will not affect the VA status nor the corresponding payoffs (agents are not pivotal). Conversely, for any profile such that $N_p = w - 1$, each non-participating agent has a unilateral incentive to deviate and pay c/w instead of the tax t. Likewise, for any profile such that $N_p > w$, each participating agent has a unilateral incentive not to participate since it will not affect the VA status but will avoid him the participation cost. Finally, all the profiles such that $N_p = w$ are PNE since any deviation of a non-participating firm will trigger tax enforcement, while non-participating agents have no interest in participating knowing that the VA is provided anyway. Consequently, there are exactly

$$|S^*| = \begin{cases} \binom{w}{N} + \sum_{k=0}^{w-2} \binom{k}{N} & \text{if } N \geq 2 \text{ and } w \geq 2 \\ \binom{w}{N} & \text{otherwise} \end{cases} \tag{8}$$

PNE in this game, none of which corresponds to the socially optimal allocation. And even though we identified PNE such that the VA succeeds, multiplicity raises the question of the VA feasibility : how will agents coordinate amongst the several subsets of Pareto equivalent PNE ?

. We conclude this preliminary study with an analysis of the symmetric MNE of the basic game. The present work focuses on both symmetric and full-support mixed strategies. It will be therefore assumed that the N agents are mixing in the support $\{0, 1\}$ according to a symmetric distribution :

$$p_i(s_i) = \begin{cases} q & \text{if } s_i = 1 \\ 1 - q & \text{if } s_i = 0 \end{cases}, \tag{9}$$

with q strictly positive. The symmetric MNE of the game are mixed strategy profiles such that the individual participation probability q is defined by

$$\binom{w-1}{N-1} q^{w-1} (1-q)^{N-w} t = \sum_{k=w-1}^{N-1} \binom{k}{N-1} q^k (1-q)^{N-1-k} \frac{C(1+k)}{1+k}, \tag{10}$$

which is the algebraic form of the condition that to contribute and not to contribute must yield the same expected gains for each agent. A general characterization[9], i.e. which also includes partial supports, is provided in Appendix A. Condition 10 allows us to establish the following proposition about the non-mediated public VA.

Proposition 1. *The public VA without the correlation device has a unique symmetric MNE, given by $Q(t)$, with $Q'(t) > 0$, $Q(0) = 0$, and $\lim_{t \to +\infty} Q(t) = 1$, where $Q(t)$ is*

[9] These results are an extension of Palfrey and Rosenthal 13 to subscription games that feature our more general payoffs structure.

defined as the inverse function of:

$$t(q) = \sum_{k=w-1}^{N-1} \frac{\binom{k}{N-1}}{\binom{w-1}{N-1}} \frac{C\,(1+k)}{1+k} \left[\frac{q}{(1-q)}\right]^{k-w+1}. \tag{11}$$

Proof 1. Algebraic manipulations lead us to rewrite (A.4) *with* $j = m = 0$ *as* $t(q)$*, which is strictly increasing in* q*:*

$$\frac{\partial}{\partial q} t(q) = \sum_{k=w-1}^{N-1} \frac{\binom{k}{N-1}}{\binom{w-1}{N-1}} \frac{C\,(1+k)}{1+k} \left[\frac{1}{(1-q)^2}\right]^{k-w+1} > 0. \tag{12}$$

Then, remark that $t(0) = 0$ *and* $\lim_{q\to 1} t(q) = +\infty$*. It follows* $t(q)$ *is invertible on our interval of interest,* $q \in [0, 1]$*.*

Finally, we state two results of interest for the efficiency analysis of section 5.

Corollary 1. *The aggregate expected payoff,* $SW\,(Q(t))$*, generated by the unique symmetric MNE of the public VA without the correlation device, is given by:*

$$SW\,(Q(t)) = -\sum_{k=w}^{N} \binom{k}{N} (Q(t))^k\,(1-Q(t))^{N-k} C(k) \tag{13a}$$

$$- \,t \sum_{k=1}^{w-1} \binom{k}{N} (Q(t))^k\,(1-Q(t))^{N-k}, \tag{13b}$$

where the sum in (13b) *corresponds to the probability of failure of the public VA.*

The second result regards the number of participating firms, which follows a binomial distribution.

Corollary 2. *The expected number of participating agents,* $E(t, N) = Q(t)N$*, increases in the threat stringency.*

4. Coordination in the mediated game

The present section is primarily devoted to study the set of equilibria of the game provided with the correlation device. This analysis allows us to state a first interesting property of mediated public VAs.

4.1. Characterization of the set of CE

We begin the analysis by introducing some useful definitions that will be used throughout the rest of this paper. First, note that provided a correlation device p on the set of strategy profiles S, one can easily derive the distributions of individual strategies s_i as the marginal distributions of a joint probability distribution. Specifically, for all i, marginal distributions are given by

$$pr(s_i = 1) = \sum_{s_{-i} \in S_{-i}} p(s_i = 1, s_{-i}) \quad \text{and} \quad pr(s_i = 0) = \sum_{s_{-i} \in S_{-i}} p(s_i = 0, s_{-i}), \tag{14}$$

with $\cup_{s_i \in S_i} (s_i \cap s_{-i})_{s_{-i} \in S_{-i}}$ an exhaustive set of events in the probability space $(S, \mathcal{P}(S))$ defined by the correlation device, where $\mathcal{P}(S)$ is the power set of S. Now, let us denote S_k the set of strategies such that the total number of participating agents is k, ie. formally

$$S_k = \{s : \sum_{i=1}^{N} s_i = k\}. \tag{15}$$

Then, note that the number of participating agents can also be defined as a random variable m, and its distribution inferred from p as follows

$$pr(m = k) = pr\left(\sum_{i=0}^{N} s_i = k\right) = \sum_{s \in S_k} p(s), \quad k \in \{0, ..., N\}, \tag{16}$$

since $\cap_{k \in \{0,...,N\}} S_k$ is also a complete system of events in $(S, \mathcal{P}(S))$. From which we get the conditional distribution of m given an individual prescription for player i, defined by

$$\sum_{k=0}^{N} pr(m = k \mid s_i) = 1 \tag{17}$$

for all $s_i \in S_i$ and for all i. Finally, let us remark that for all $k \in \{0, ..., N\}$, the set S_k can be partitioned according to the value of s_i, with

$$pr\,(m = k) = pr\,(m = k \cap s_i = 0) + pr\,(m = k \cap s_i = 1)\,, \tag{18}$$

where the associated conditional distribution of s_i given a number of participants $m = k$ is

$$\sum_{s_i \in \{0,1\}} pr(s_i \mid m = k) = 1. \tag{19}$$

Still following Myerson 10 interim definition of correlated equilibria, we can write the strategic incentive constraints in the general case. When agents are privately told to play $s_i = 0$, the distribution p that was prealably announced by the regulator must be such that they have no incentive to unilateraly deviate, ie.

$$-\sum_{k=0}^{k \le w-1} pr\,(m = k \mid s_i = 0)\, t - \sum_{k=w}^{N-1} pr\,(m = k \mid s_i = 0)\, 0 \ge -\sum_{k=0}^{k \le w-2} pr\,(m = k \mid s_i = 0)\, t \\ - \sum_{k=w-1}^{N-1} pr\,(m = k \mid s_i = 0)\, \frac{C(k+1)}{k+1}$$

must holds for all i, which simplifies to:

$$pr\left(m=w-1 \mid s_i=0\right)\left(t-\frac{c}{w}\right)-\sum_{k=w}^{N-1} pr\left(m=k \mid s_i=0\right)\frac{C(k+1)}{k+1} \leq 0. \tag{20}$$

Likewise, when the prescription is $s_i = 1$,

$$-\sum_{k=1}^{k\leq w-1} pr\left(m=k \mid s_i=1\right)t-\sum_{k=w}^{N} pr\left(m=k \mid s_i=1\right)\frac{C(k)}{k} \geq -\sum_{k=1}^{k\leq w} pr\left(m=k \mid s_i=1\right)t - \sum_{k=w+1}^{N} pr\left(m=k \mid s_i=1\right)0$$

must hold for all i, which simplifies to:

$$pr\left(m=w \mid s_i=1\right)\left(t-\frac{c}{w}\right)-\sum_{k=w+1}^{N} pr\left(m=k \mid s_i=1\right)\frac{C(k)}{k} \geq 0. \tag{21}$$

Finally, the probability constraints are given by

$$\begin{cases} \sum_{k=0}^{N} pr\left(m=k\right)=1 \\ \left(p_{(0)_N},...,p_{(1)_N}\right)\in\mathbb{R}^+ \end{cases}. \tag{22}$$

Conditions (20) (21), (22) fully characterize the set of CE of the participation game.

4.2. Coordination on the symmetric MNE

In order to describe the relationship between the basic and mediated public VAs' outcomes, we first need to introduce a last crucial element of terminology.

Definition 1. *A correlated strategy $p \in \triangle(S)$ is* symmetrical *if and only if,*

$$p(s)=\frac{pr(m=k)}{\binom{k}{N}}, \qquad \forall s\in S_k, \qquad k\in[0,N], \tag{23}$$

i.e. participation profiles with the same number of participating firms are equiprobable.

Note that symmetry implies the marginal distributions verify:

$$pr(s_i=1)=\sum_{s_{-i}\in S_{-i}} p(s_i=1,s_{-i})=q \tag{24a}$$

$$pr(s_i=0)=\sum_{s_{-i}\in S_{-i}} p(s_i=0,s_{-i})=1-q, \tag{24b}$$

for all i, with $q \in [0, 1]$. Now suppose there exists some marginal distributions $(q, 1-q)$ such that (10) holds. It can be shown that the induced distribution on S also verifies (20) and (21). Specifically, it belongs to the subset of symmetrical EC, and uniquely satisfies an extra-assumption of individual participation decisions independence. This result allows us to derive a first practical indication about public VAs policies.

Proposition 2. *The symmetric MNE of the public VA without a correlation device can be implemented by the regulator by using the correlation device.*

Proof 2. *See Appendix B.*

In other words, adding the correlation device allows to overcome the coordination issues that may arise when a public VA is implemented.

5. Efficiency analysis and impact of the tax threat

In this section we show that beyond solving the coordination issue, the correlation device may also allow to improve collective welfare, for all t. We then turn to an analysis of the impact of the tax threat on the set of reachable payoff profiles.

5.1. The optimal CE

Consider a regulator using the public VA provided with the correlation device. He seeks to implement some distribution p on the set of strategy profiles which maximizes:

$$\max_{p\in\triangle(S)} -Nt\sum_{k=0}^{w-1} pr\left(m=k\right) - pr\left(m=w\right)C(w) - \sum_{k=w+1}^{N} pr\left(m=k\right)C(k) \tag{25}$$

subject to the constraints (20), (21) and (22) . Let us restrict to symmetrical distributions such as specified in definition 1. Remark, then, that this assumption implies in particular:

$$pr\left(m=k \mid s_i=1\right) = \frac{\binom{k}{N-1}}{\binom{k-1}{N-1}} pr\left(m=k \mid s_i=0\right) = \frac{k}{N} pr\left(m=k\right) \quad \forall k\in]0,N[. \tag{26}$$

Substituting (26) into both the objective and the constraints, the latter conveniently rewrites:

<u>s.t</u>

$$pr(m=w-1 \mid s_i=1)\frac{N-w+1}{w-1}\left(\frac{C(w)}{w}-t\right)+\sum_{k=w}^{N-1}\frac{N-k}{k}pr(m=k \mid s_i=1)\frac{C(k+1)}{k+1}\geq 0 \tag{27a}$$

$$pr\left(m=w \mid s_i=1\right)\left(t-\frac{C(w)}{w}\right)-\sum_{k=w+1}^{N} pr\left(m=k \mid s_i=1\right)\frac{C\left(k\right)}{k}\geq 0 \tag{27b}$$

$$pr(m=0 \mid s_i=0)+pr\left(m=N \mid s_i=1\right)+\sum_{k=1}^{N-1}\left(\frac{N}{k}pr\left(m=k \mid s_i=1\right)\right)=1 \tag{27c}$$

We recognize a linear programming problem, which can be converted into an augmented form by introducing two slack variables, denoted x_1 (into constraint (27a)), and x_2 (into constraint (27b)). Note that the standard program generated is composed of $j = 3$ equality constraints, and $l = N + 3$ variables only. Indeed, the symmetry hypothesis implies the objective is actually optimized in $\{pr(m = k)\}_0^N$, conditionally to some prescription that we arbitrarily choose to be $s_i = 1$. It can be solved by applying the two-phase simplex algorithm[10]**?** . However, such a method may imply numerous (and cumbersome) iterations if it is started without proceeding to a preliminary heuristic analysis. The next proposition states the result of the program, while the proof provides both an intuitive and a formal (Appendix C) argument.

Proposition 3. *The symmetrical CE maximizing the aggregate payoff* (25) *is given by* p^**, the unique symmetrical probability distribution on* S *which is induced by:*

$$pr^*(m = N) = \frac{(tw - C(w))}{(tw - C(w)) + C(N)}, \tag{28a}$$

$$pr^*(m = w) = \frac{C(N)}{(tw - C(w)) + C(N)}, \tag{28b}$$

and $pr^*(m = k) = 0$*, for all* $k \in [0, N] \setminus \{w, N\}$*.*

Proof 3. First, observe that (27a) *and* (27b) *do not depend either on* $pr(m = 0)$ *or* $\{pr(m = k \mid s_i = 1)\}_1^{w-2}$*. Since* $t > c$*, we know the optimal distribution must assign a probability* 0 *to the corresponding participation profiles. Then, remark that* $pr(m = w-1 \mid s_i = 1)$ *appears in* (27a) *solely, and that it does not need to be strictly positive. However, given that* $pr(m = 0)$ *and* $\{pr(m = k \mid s_i = 1)\}_1^{w-2}$ *are set to* 0*, the probability* $pr\,(m = w \mid s_i = 1)$ *must now be strictly positive for* (27b) *to hold. It follows* (27a) *cannot be binding at the optimum. Finally, both the assumptions on costs and the structure of* (27b) *induce it will be binding since it is now straightforward the highest probability should be put on profile* $(1)_N$*. As a result, we know* $\{x_1,\ pr(m{=}w|s_i{=}1),\ pr(m = N)\}$ *is the basis which has to be tested in order to confirm our heuristic solution is optimal. This is done in Appendix C.*

Now that we have characterized the optimal CE of the game, let us state some normative consequences on public VAs emerging from a background threat.

Corollary 3. *With the optimal correlation device, the VA always succeeds and the aggregate expected payoff is given by:*

$$SW^*_{EC}(t) = \frac{C(N)\,(c(w) - tw)}{(tw - C(w)) + C(N)}. \tag{29}$$

Even though our specifications are too general for $p^*(t)$ and $(Q(t), 1{-}Q(t))$ to be explicitly ranked in terms of social welfare, since it has been shown in proposition 3 that the unique SMNE of the basic game is also a CE of the mediated game, the following can be stated:

10

Corollary 4. *$SW^*_{EC}(t) \geq SW(Q(t))$, for all t.*

Let us stress that the inequality holds for all t, which implies it holds, in particular, for the level of tax threat that would be credible given some further step of the game. For instance, such a level may correspond, as in the specified example of section 6, to a tax on emissions such that E is reached under the mandatory scheme for subgame perfection.

5.2. Impact of the tax threat

Corollary 5. *Under the symmetric MNE, an increase of the tax threat leads to an increase of both the expected number of participating agents, $E(t,N)$, and the collective welfare, $SW(Q(t))$.*

Proof 4. Observe that:

$$\frac{\partial}{\partial q} pr(m=k) = kQ^{k-1}(1-Q)^{N-k} - (N-k)(1-Q)^{N-k-1},$$

hence we know the probability on profiles with k participating firms increases in Q if and only if $Q > {}^{k}/_{N}$ or, equivalently, $k > E(t,N)$, where

$$\frac{\partial}{\partial t} E(t,N) = N \frac{\partial}{\partial t} Q(t) > 0.$$

Moreover, the cost of the public VA is decreasing, by assumption, in the number of participating agents. It follows that an increase of $pr(m=k)$ for all $k > E(t,N)$ implies the expected aggregate payoff increases as well.

We now focus of on the public VA provided with the correlation device specifically. In order to study the impact of the tax threat on the set of CE, let us rewrite (27a) and (27b) as follows:

$$\sum_{k=w+1}^{N} \frac{N-(k-1)}{N-(w-1)} \left(\frac{pr(m=k-1)}{pr(m=w-1)}\right) \frac{C(k)}{k} \geq t - \frac{C(w)}{w} \geq \sum_{k=w+1}^{N} \frac{k}{w} \left(\frac{pr(m=k)}{pr(m=w)}\right) \frac{C(k)}{k}, \tag{30}$$

where each ratio of probabilities can be interpreted as a deviation's pivotality rate. Now, remark that a more stringent tax does not impact the RHS inequality, i.e. the unilateral incentives to comply when the regulator recommends $s_i = 1$. Indeed, if there was already no incentive not to participate for some t, then a firm will be even less willing to take the risk the public VA fails when the threat is higher. However, an increase of t does directly impact the LHS, which can be seen as what the firm would gain by not participating when $s_i = 0$ is prescribed, and which must therefore remain smaller than the tax level so the firm does not choose to deviate by participating. Thus, one effect generated by a higher threat is that distributions with higher probabilities on less efficient profiles become equilibria of the mediated game. This result is consistent with Chiambretto and Stahn 7 and Suter et al. 15, in which it is shown that an increase of the pigovian tax level in an emission game leads to under-participation, and thus inefficiency under their assumption, in a preemptive public VA.

We now turn to studying the impact of the tax threat on the optimal CE specifically.

Proposition 4. *Under the public VA provided with the correlation device, a higher threat rises the probability on the socially optimal outcome.*

Proof 5. Under our assumptions:

$$\frac{\partial}{\partial t} pr^*(m = N) = \frac{wC(N)}{(tw - C(w) + C(N))^2} > 0,$$

i.e. the probability assigned to the full participation profile by p^* strictly increases in t.

Proposition 5. *Collective welfare associated with the mediated public VA increases in tax threat stringency. Specifically, the expected aggregate payoff tends toward the level generated by the full participation profile, $-C(N)$.*

Proof 6. Under our assumptions:

$$\frac{\partial}{\partial t} SW^*(t) = \frac{-wC(N) - wC(N)(C(w) - tw)}{(tw - C(w) + C(N))^2} > 0,$$

i.e. the social aggregate payoffs under the optimal correlated equilibrium strictly increases in t, and

$$\lim_{t \to +\infty} \frac{(tw - C(w))}{(tw - C(w)) + C(N)} = 1.$$

6. A numerical example

We apply our game of participation to a public VA to a specified pollution abatement model and calculate, for different parameters values, the corresponding unique MNE and optimal CE, as functions of the tax threat.

6.1. The pollution abatement model

Consider N symmetric firms producing a good and pollutant emissions which are engaged in Cournot competition. Let $\pi(e_i)_{i=1}^{N}$ be the indirect profit function with ${}^{\partial^2 \pi(e_i)}/_{\partial^2 e_i} \leq 0$. Specifically, assume $\pi(e_i) = 20e_i - e_i^2$, with $e_i^* = 10$ the optimal level of emission at laisser faire. Let us denote $E \equiv N(e_i^* - \epsilon)$ the emission target, which amounts to an aggregate reduction of $N\epsilon$ emissions units. Assume furthermore the regulator chooses w such that $(N - w) e_i^* = N(e_i^* - \epsilon)$, i.e. $w = {}^{N\epsilon}/_{10}$. As in the general participation game, let us denote t the tax threat. The payoffs of firm i in the basic game are given by :

$$u_i(s) = \begin{cases} -\left(\frac{\epsilon N}{N_p}\right)^2 & s_i = 1 \text{ and } N_p \geq w \\ 0 & s_i = 0 \text{ and } N_p \geq w \\ -t & N_p < w \end{cases} \tag{31}$$

where ${}^{C(N_p)}/{}_{N_p} = \left({}^{\epsilon N}/{}_{N_p}\right)^2$ is the individual participation cost which is decreasing in N_p from individual profits' concavity. To be more specific:

$$-\frac{C\left(N_p\right)}{N_p} = -\left(\pi\left(e_i^*\right) - \pi\left(\frac{N\left(e_i^* - \epsilon\right) - \left(N - N_p\right)e_i^*}{N_p}\right)\right) = -\left(\epsilon\frac{N}{N_p}\right)^2.$$

This illustration also requires we specify what would be the tax threat in such a specified context. We assume t is the difference between the profit at laissez-faire and the profit under the mandatory regulation scheme. Let us say the mandatory regulation scheme is two-part. Specifically, it is primarily composed of a pigovian tax, i.e. a tax on emissions set at a level t^{PIG} such that each of the N firms produces $e_i = \left(e_i^{LF} - \epsilon\right)$ under the mandatory regulation scheme, and the target E is reached:

$$t^{PIG} = \left.\frac{\partial\pi\left(e_i\right)}{\partial e_i}\right|_{e_i^{LF}-\epsilon} = 2\epsilon. \tag{32}$$

The second part of the mandatory regulation scheme is a lump-sum transfer, which may be interpreted as transaction costs, denoted TC, and such that :

$$\begin{aligned} t &= \pi_i(e_i^*) - \pi_i\left(e_i^* - \epsilon\right) + t^{PIG}\left(e_i^* - \epsilon\right) + TC \\ &= \epsilon^2 + (10-\epsilon)t^{PIG} + TC. \end{aligned}$$

We now calculate the optimal CE, the unique SME and the social welfare associated with the optimal CE, as functions of N, w and the target ϵ.

6.2. Non-mediated and mediated game's equilibria:

The optimal CE of the game is given by:

$$pr^*(m=N) = \frac{\left(\left(\epsilon^2 + (10-\epsilon)t^{PIG} + TC\right)w - \left(\frac{(\epsilon N)^2}{w}\right)\right)}{\left(\left(\epsilon^2 + (10-\epsilon)t^{PIG} + TC\right)w - \left(\frac{(\epsilon N)^2}{w}\right)\right) + \left(\frac{(\epsilon N)^2}{N}\right)},$$

$$pr^*(m=w) = \frac{\left(\frac{(\epsilon N)^2}{N}\right)}{\left(\left(\epsilon^2 + (10-\epsilon)t^{PIG} + TC\right)w - \left(\frac{(\epsilon N)^2}{w}\right)\right) + \left(\frac{(\epsilon N)^2}{N}\right)},$$

and $pr^*(m=k) = 0$, for all $k \in [0,N] \setminus \{w,N\}$. For a comparison purpose, we also calculate the induced marginal distributions:

$$\begin{aligned} pr(s_i=1) &= pr\left(s_i = 1 \cap m = N\right) + pr\left(s_i = 1 \cap m = w\right) \\ &= \frac{\left(\left(\epsilon^2 + (10-\epsilon)t^{PIG} + TC\right)w - \left(\frac{(\epsilon N)^2}{w}\right)\right) + w\epsilon^2}{\left(\left(\epsilon^2 + (10-\epsilon)t^{PIG} + TC\right)w - \left(\frac{(\epsilon N)^2}{w}\right)\right) + \left(\frac{(\epsilon N)^2}{N}\right)}, \end{aligned}$$

$$pr(s_i = 0) = pr\left(s_i = 0 \cap m = N\right) + pr\left(s_i = 0 \cap m = w\right)$$
$$= \frac{(N-w)\,\epsilon^2}{\left(\left(\epsilon^2 + (10-\epsilon)t^{PIG} + TC\right)w - \left(\frac{(\epsilon N)^2}{w}\right)\right) + \left(\frac{(\epsilon N)^2}{N}\right)},$$

while the expected social welfare is given by:

$$-\frac{\left(\frac{(\epsilon N)^2}{N}\right)\left(\left(\epsilon^2 + (10-\epsilon)t^{PIG} + TC\right)w\right)}{\left(\left(\epsilon^2 + (10-\epsilon)t^{PIG} + TC\right)w - \left(\frac{(\epsilon N)^2}{w}\right)\right) + \left(\frac{(\epsilon N)^2}{N}\right)}.$$

Finally, we characterize the symmetric MNE of the game without the correlation device:

$$t(q) = \sum_{w-1}^{N-1} \left(\frac{(w-1)!\,(N-w)!}{k!\,(N-1-k)!}\right)\left(\frac{\epsilon N}{k+1}\right)^2 \left(\frac{q}{1-q}\right)^{k-w+1}$$

6.3. The numerical results

Our results are summarized in figures 1, 2 and 3. Specifically, Figure 1 draws the individual participation probability induced by the optimal CE under the mediated public VA policy as a function of the tax threat. The blue line corresponds to the vector of parameters values $(N, \epsilon, w) = (5, 4, 2)$, while the red and the black lines correspond to vectors $(10, 6, 6)$ and $(10, 2, 2)$ respectively. Figure 2 simply draws $t(q)$ for the same three vectors of parameters values.

Figure 3 focuses on the case $(N, \epsilon, w) = (5, 4, 2)$. It shows on the same graph both $t(q)$ which is generated by the non-mediated game (black line), and the tax-threat level necessary to reach some participation rate under the optimal CE of the mediated game (blue line).

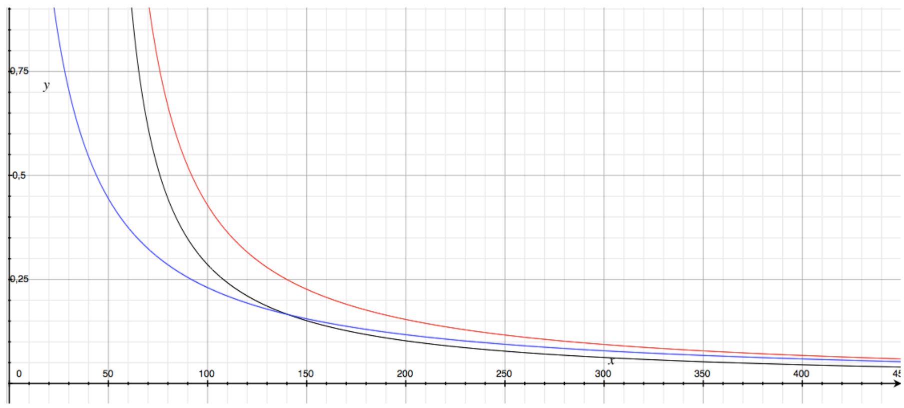

Figure 1: $Pr\,(s_i = 1)$ for several values of parameters (N, ϵ, w). blue: $N = 5, \epsilon = 4, w = 2$; red : $N = 10, \epsilon = 6, w = 6$; and black : $N = 10, \epsilon = 2, w = 2$.

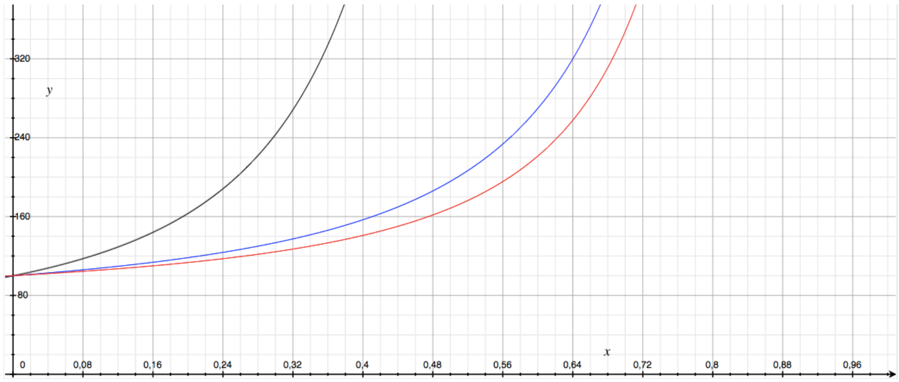

Figure 2: $t(q)$ for several values of parameters (N, ϵ, w). blue: $N = 5, \epsilon = 4, w = 2$; red : $N = 10, \epsilon = 6, w = 6$; and black : $N = 10, \epsilon = 2, w = 2$.

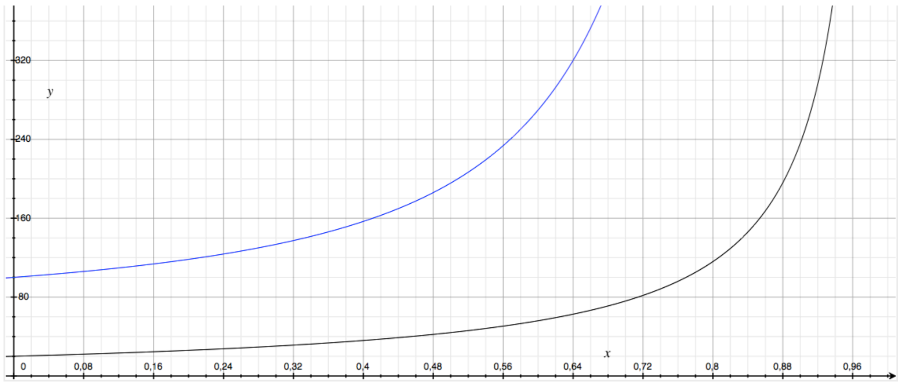

Figure 3: $t(q)$ (black) and $(Pr\,(s_i = 1))^{-1}(t)$ (blue), for $(N, \epsilon, w) = (5, 4, 2)$.

7. Concluding remarks

We have developed a subscription game with a payoff structure representing agents' participation incentives under a preemptive public VA policy. Two key features of such a mechanism are the exogenous global target and the use of a tax threat with a collective liability. The first feature is represented by the minimum participation threshold w, while the second feature has lead us to modify the general subscription game by adding a punishment parameter t, that applies to strategy profiles such that the public good is not provided or, in other words, when the public VA fails.

All our results are demonstrated in the general N-player case. We characterize the set of correlated equilibria and thereby the circumstances under which the public VA provided with a correlation device may succeed. Specifically, we characterize the unique symmetric MNE of the game and show that it can be implemented by the regulator by using the correlation device. Then, we find that the device not only solve the problem raised by multiplicity, but also ensures efficiency gains. Indeed, we characterize the optimal CE of the mediated public VA, with the related finding that the unique symmetric MNE yields a smaller aggregate payoff, for all t (we know a credibility requirement would actually limit the regulator in his choice of t). Finally, we find that a higher threat rises the probability on the socially optimal allocation and improves collective welfare for both equilibria.

Two research directions are left for further work. First, the numerical example provided in the last section reveals there is some potential for considering asymmetric firms in this framework. We furthermore notice that, as in Brau and Carraro 4, it could be interesting to check if the analysis is robust to the existence of partial spillover (i.e. rewrite the payment matrix assuming non participating agents enjoy a share $\alpha < 1$ of the regulatory gains enjoyed by participating agents).

Appendix

Appendix A. Mixed strategies Nash equilibria of the basic game

As in Palfrey and Rosenthal 13, we restrict our analysis to the cases such that there are j agents with a strategy support $\{0\}$ and m agents with a strategy support $\{1\}$. The rest of agents is mixing in the support $\{0, 1\}$ according to a symmetrical distribution :

$$p_i(s_i) = \begin{cases} q & \text{if } s_i = 1 \\ 1-q & \text{if } s_i = 0 \end{cases}. \tag{A.1}$$

with q strictly positive, and (m, j, w, N) an admissible vector of parameters such as defined in Palfrey 1984 (ref.), namely

$$(m, j, w, N) \in \{(m, j, w, N) \mid 0 \leq j \leq N - w \text{ and } 0 \leq m \leq w - 1\}.$$

Admissibility both guarantees that the parameters define a partition of the set of players (ie. $N-m-j \geq 0$), and that there exists a unique and strictly positive best response probability q for mixing agents. Indeed, if $j > N - w$, mixing agents are not pivotal and any $q \in [0, 1]$ is a best response. Likewise, if $m > w - 1$, the unique best response of potentially mixing agents is $q = 0$. For the sake of convenience, let us introduce the notation

$$A = N - m - j \tag{A.2}$$

$$B = w - m. \tag{A.3}$$

before seeking for admissible combinations of parameters values (w, m, j, q) and N, such that no player has an incentive to unilaterally change his strategy in the model. It is the case when the following incentive constraints are simultaneously satisfied

$$\begin{aligned} -\binom{m}{m+j} \sum_{k=B-1}^{A-1} \binom{k}{A-1} q^k (1-q)^{A-1-k} \frac{C(1+m+k)}{1+m+k} \\ -\binom{m}{m+j} \sum_{k=0}^{B-2} \binom{k}{A-1} q^k (1-q)^{A-1-k} t = \\ -t\binom{m}{m+j} \sum_{k=0}^{B-1} \binom{k}{A-1} q^k (1-q)^{A-1-k} \end{aligned}$$

$$\Leftrightarrow \binom{B-1}{A-1} q^{B-1} (1-q)^{A-B} t = \sum_{k=B-1}^{A-1} \binom{k}{A-1} q^k (1-q)^{A-1-k} \frac{C(1+m+k)}{1+m+k}, \tag{A.4}$$

which is the algebraic form of the condition that to contribute and not to contribute must yield the same expected gains for the mixing agents, and

$$-\binom{m-1}{m+j-1}\sum_{k=B}^{A}\binom{k}{A}q^{k}(1-q)^{A-k}\frac{C(m+k)}{m+k}$$
$$-\binom{m-1}{m+j-1}\sum_{k=0}^{B-1}\binom{k}{A}q^{k}(1-q)^{A-k}t \geq$$
$$\binom{m-1}{m+j-1}\sum_{k=0}^{B}\binom{k}{A}q^{k}(1-q)^{A-k}t$$
$$\Leftrightarrow \binom{B}{A}q^{B}(1-q)^{A-B}t \geq \sum_{k=B}^{A}\binom{k}{A}q^{k}(1-q)^{A-k}\frac{C(m+k)}{m+k}, \tag{A.5}$$

$$-\binom{m}{m+j-1}\sum_{k=0}^{B-1}\binom{k}{A}q^{k}(1-q)^{A-k}t \geq$$
$$-\binom{m}{m+j-1}\sum_{k=B-1}^{A}\binom{k}{A}q^{k}(1-q)^{A-k}\left(\frac{C(m+k)}{m+k}\right)$$
$$-\binom{m}{m+j-1}\sum_{k=0}^{B-2}\binom{k}{A}q^{k}(1-q)^{A-k}t$$
$$\Leftrightarrow \binom{B-1}{A}q^{B-1}(1-q)^{A-B+1}t \leq \sum_{k=B-1}^{A}\binom{k}{A}q^{k}(1-q)^{A-k}\left(\frac{C(m+k)}{m+k}\right), \tag{A.6}$$

which are the algebraic forms for the conditions that (i) contributing is at least as good than not contributing for participating agents (A.5) and (ii) not contributing is at least as good than contributing for non participating agents (A.6). These results are an extension of Palfrey and Rosenthal 13 to subscription games with our more general payoffs structure.

Note that if (i) $m=0$, only conditions (A.4) and (A.6) need to be satisfied (ii) $j=0$, only conditions (A.5) and (A.6) need to be satisfied (iii) $j=m=0$, only condition (A.6) applies and (iv) for $m+j=N-1$, the admissibility constraints hold at equality and the conditions rewrite as the pure Nash equilibria conditions with $q\in\{0,1\}$.

Appendix B. Proof of Proposition 2

Let us remark that symmetry, added to the independence of individual participation decisions, imply:

$$pr\,(m=k\mid s_i=1)=\frac{pr\,(m=k)\,pr\,(s_i=1\mid m=k)}{pr\,(s_i=1)}=\left(\binom{N}{k}q^{k}(1-q)^{N-k}\right)\frac{k}{qN}$$
$$pr\,(m=k\mid s_i=0)=\frac{pr\,(m=k)\,pr\,(s_i=0\mid m=k)}{pr\,(s_i=0)}=\left(\binom{N}{k}q^{k}(1-q)^{N-k}\right)\frac{N-k}{N}\frac{1}{1-q},$$

Now, using the two previous equalities, we can rewrite condition (20) of EC as follows:

$$\binom{N}{w-1}q^{w-1}(1-q)^{N-w+1}\frac{N-w+1}{N(1-q)}\left(t-\frac{c}{w}\right)-\sum_{k=w}^{N-1}\binom{N}{k}q^{k}(1-q)^{N-k}\frac{N-k}{N(1-q)}\frac{C(k+1)}{k+1}\leq 0$$

$$\binom{N}{w-1}q^{w-1}(1-q)^{N-w+1}\frac{N-w+1}{N(1-q)}\left(t-\frac{c}{w}\right)-\frac{1}{(1-q)\,N}\sum_{k=w}^{N-1}\binom{N}{k+1}q^{k}(1-q)^{N-k}C(k+1)\leq 0$$

$$(N-w+1)\binom{N}{w-1}q^{w-1}(1-q)^{N-w+1}\left(t-\frac{c}{w}\right)-\sum_{k=w+1}^{N}\binom{N}{k}q^{k-1}(1-q)^{N-k+1}C(k)\leq 0 \tag{B.1}$$

Likewise, condition (21) of EC rewrites:

$$\left(\binom{N}{w}q^{w}(1-q)^{N-w}\right)\frac{w}{qN}\left(t-\frac{c}{w}\right)+\sum_{k=w+1}^{N}\left(\binom{N}{k}q^{k}(1-q)^{N-k}\right)\frac{k}{qN}\left(-\frac{C(k)}{k}\right)\geq 0$$

$$\left(\binom{N}{w}q^{w}(1-q)^{N-w}\right)\frac{w}{qN}\left(t-\frac{c}{w}\right)-\frac{1}{(1-q)\,N}\sum_{k=w+1}^{N}\left(\binom{N}{k}q^{k-1}(1-q)^{N-k+1}\right)C(k)\geq 0$$

$$(N-w+1)\binom{N}{w-1}q^{w-1}(1-q)^{N-w+1}\left(t-\frac{c}{w}\right)-\sum_{k=w+1}^{N}\binom{N}{k}q^{k-1}(1-q)^{N-k+1}C(k)\geq 0 \tag{B.2}$$

Now, notice that (B.1) and (B.2) imply:

$$(N-w+1)\left(\binom{N}{w-1}q^{w-1}(1-q)^{N-w+1}\right)\left(t-\frac{c}{w}\right)=\sum_{k=w+1}^{N}\left(\binom{N}{k}q^{k-1}(1-q)^{N-k+1}\right)C(k).$$

$$\Leftrightarrow (N-w+1)\binom{N}{w-1}q^{w-1}(1-q)^{N-w+1}\left(t-\frac{c}{w}\right)=N\sum_{k=w}^{N-1}\binom{N-1}{k}q^{k}(1-q)^{N-k}\frac{C(k+1)}{k+1}$$

$$\Leftrightarrow \binom{N-1}{w-1}q^{w-1}(1-q)^{N-w+1}\left(t-\frac{c}{w}\right)=\sum_{k=w}^{N-1}\binom{N-1}{k}q^{k}(1-q)^{N-k}\frac{C(k+1)}{k+1},$$

which is the symmetrical ENSM condition when m=j=0.

Appendix C. Proof of Proposition 3

Let us write the standard form constraints, and rearrange them so as to isolate the basis variables x_1, $pr(m=w\mid s_i=1)$ and $pr(m=N)$:

$$\begin{aligned}x_1 &= \left(\frac{N-w+1}{w-1}\frac{C(w)-tw}{w}\right)pr\left(m=w-1\mid s_i=1\right)\\ &+\sum_{k=w+1}^{N-1}\left(\frac{N-k}{k}\frac{C\left(k+1\right)}{k+1}-\left(\left(\frac{C\left(w+1\right)}{w+1}\frac{\left(N-w\right)}{C(w)-tw}\right)\frac{C\left(k\right)}{k}\right)\right)pr\left(m=k\mid s_i=1\right)\\ &-\left(\left(\frac{C\left(w+1\right)}{w+1}\right)\frac{N-w}{C(w)-tw}\right)\frac{C\left(N\right)}{N}pr\left(m=N\mid s_i=1\right)-\left(\left(\frac{C\left(w+1\right)}{w+1}\right)\frac{N-w}{C(w)-tw}\right)x_2\end{aligned} \tag{C.1}$$

$$\begin{aligned}pr\left(m=N\mid s_i=1\right) &= \left[1+\frac{C\left(N\right)}{C(w)-tw-C\left(N\right)}\right]+\left(\frac{N}{C(w)-tw-C\left(N\right)}\right)x_2\\ &-\left(1+\frac{C\left(N\right)}{C(w)-tw-C\left(N\right)}\right)pr\left(m=0\mid s_i=0\right)\\ &-\sum_{k=1}^{w-1}\frac{N}{k}\left(\frac{C\left(N\right)}{C(w)-tw-C\left(N\right)}+1\right)pr\left(m=k\mid s_i=1\right)\\ &-\sum_{k=w+1}^{N-1}\frac{N}{k}\left(1+\frac{\left(C\left(N\right)-C\left(k\right)\right)}{C(w)-tw-C\left(N\right)}\right)pr\left(m=k\mid s_i=1\right)\end{aligned} \tag{C.2}$$

$$
\begin{aligned}
pr\left(m=w \mid s_i=1\right) = & -\left(\frac{w}{C(w)-tw-C\left(N\right)}\right)x_2-\left(\frac{C\left(N\right)}{C(w)-tw-C\left(N\right)}\right)\frac{w}{N} \\
& -\sum_{k=w+1}^{N-1}\frac{w}{k}\left(\frac{C\left(k\right)-C\left(N\right)}{C(w)-tw-C\left(N\right)}\right)pr\left(m=k \mid s_i=1\right) \\
& +\left(\left(\frac{C(N)}{C(w)-tw-C\left(N\right)}\right)\frac{w}{N}\right)pr\left(m=0 \mid s_i=0\right) \\
& +\sum_{k=1}^{w-1}\frac{w}{k}\left(\frac{C(N)}{C(w)-tw-C\left(N\right)}\right)pr\left(m=k \mid s_i=1\right)
\end{aligned} \tag{C.3}
$$

We substitute (C.2) and (C.3) into the objective (25) after it was rewritten according to the symmetry assumption (26). It follows the objective depends only on the N non-basic variables :

$$
C(w)\left(\frac{C\left(N\right)}{C(w)-tw-C\left(N\right)}\right)-C(N)\left(\frac{C(w)-tw}{C(w)-tw-C\left(N\right)}\right)+x_2N\left[\frac{C(w)-C(N)}{C(w)-tw-C\left(N\right)}\right] \tag{C.4a}
$$

$$
+pr\left(m=0 \mid s_i=0\right)\left[C(N)\left(\frac{C(w)-tw}{C(w)-tw-C\left(N\right)}\right)-C(w)\left(\frac{C\left(N\right)}{C(w)-tw-C\left(N\right)}\right)-Nt\right] \tag{C.4b}
$$

$$
+\sum_{k=1}^{w-1}pr\left(m=k \mid s_i=1\right)\frac{N}{k}\left[C(N)\left(\frac{C(w)-tw}{C(w)-tw-C\left(N\right)}\right)-C(w)\left(\frac{C\left(N\right)}{C(w)-tw-C\left(N\right)}\right)-Nt\right] \tag{C.4c}
$$

$$
+\sum_{k=w+1}^{N-1}pr\left(m=k \mid s_i=1\right)\frac{N}{k}\left[C(N)\left(\frac{C(w)-tw-C\left(k\right)}{C(w)-tw-C\left(N\right)}\right)-C(w)\left[\frac{\left(C\left(N\right)-C\left(k\right)\right)}{C(w)-tw-C\left(N\right)}\right]-C(k)\right] \tag{C.4d}
$$

Then, let us remark that

$$
C(N)\left(\frac{C(w)-tw}{C(w)-tw-C\left(N\right)}\right)-C(w)\left(\frac{C\left(N\right)}{C(w)-tw-C\left(N\right)}\right)=\left(\frac{-C(N)\left(tw\right)}{C(w)-tw-C\left(N\right)}\right)>0 \tag{C.5}
$$

since $t>c$, and that $\left(\frac{C(N)}{C(N)+tw-C(w)}\right)<1$, from which it follows:

$$
\left(\frac{-C(N)\left(tw\right)}{C(w)-tw-C\left(N\right)}\right)<Nt. \tag{C.6}
$$

Coefficients (dual variables) associated to $\{pr(m=0),\{pr(m=k \mid s_i=1)\}_0^{w-1}\}$ are therefore negative. Besides, we have that

$$
C(N)+C(N)\frac{\left(C(N)-C\left(k\right)\right)}{C(w)-tw-C\left(N\right)}<C(k)+C(w)\left[\frac{\left(C\left(N\right)-C\left(k\right)\right)}{C(w)-tw-C\left(N\right)}\right], \tag{C.7}
$$

it follows

$$
C(N)\underbrace{\left(\frac{C(w)-tw-C\left(k\right)}{C(w)-tw-C\left(N\right)}\right)}_{>0}-C(w)\underbrace{\left(\frac{\left(C\left(N\right)-C\left(k\right)\right)}{C(w)-tw-C\left(N\right)}\right)}_{>0}-C(k)<0. \tag{C.8}
$$

Therefore, coefficients (dual variables) associated to $\{pr(m=k \mid s_i=1)\}_{w+1}^{N-1}$ are also negative. Finally, it is straightforward the dual variable of x_2 in (C.4a) is negative as well. Since non-basic variables are all set to 0 while constrained to non-negativity, it can be concluded the program is solved for:

$$\begin{cases} \frac{N-w}{w} pr\left(m=w \mid s_i=1\right) \frac{C(w+1)}{w+1} = x_1 & \forall i \\ pr\left(m=w \mid s_i=1\right)\left(t-\frac{C(w)}{w}\right) - pr\left(m=N \mid s_i=1\right)\frac{C(N)}{N} = 0 & \forall i \\ pr\left(m=w \mid s_i=1\right) + pr\left(m=N \mid s_i=1\right) = 1 \end{cases}$$

$$\Rightarrow \begin{cases} x_1^* = \frac{N-w}{w}\left(\frac{(tw-C(w))}{(tw-C(w))+C(N)}\right)\frac{C(w+1)}{w+1} \\ pr^*(m=w \mid s_i=1) = \frac{w}{N}\frac{C(N)}{(tw-C(w))+C(N)} \\ pr^*(m=N) = \frac{(tw-C(w))}{(tw-C(w))+C(N)}, \end{cases} \tag{C.9}$$

from which we get: $pr^*(m=w \mid s_i=0) = \frac{N-w}{N}\frac{C(N)}{(tw-C(w))+C(N)}$.

Appendix D. The two-player example

In order to illustrate the general study, let us consider the game when $N=2$ and $w=1$. The strategic form is the following :

	$s_2=1$	$s_2=0$
$s_1=1$	$-\frac{C(2)}{2}, -\frac{C(2)}{2}$	$-c, 0$
$s_1=0$	$0, -c$	$-t, -t$

where $C(w)=C(1)\equiv c$. From the study above, we know this game has two pure Nash equilibria corresponding to the strategy profiles such that $m=w$, namely $(1,0)$ and $(0,1)$, and that the set of admissible vectors is

$$(m,j,w,N) = \{(0,0,1,2), (0,1,1,2)\}.$$

Substituting the parameters values in the relevant general conditions (A.6) and (A.4), we get

$$\begin{cases} (1-q)\,t = (1-q)\,c + q\frac{C(2)}{2} & \text{if } j=0 \\ t=c \text{ and } 0 \leq qc & \text{if } j=1 \end{cases}, \tag{D.1}$$

conjointly characterizing the symmetrical mixed Nash equilibria of the game. Note then that the inequalities induced by $j=1$ imply that mixed Nash equilibria with asymmetrical supports exist if and only if $t=c$, in which case one player i does not contribute while any distribution on S_{-i} is also a best response for its opponent. Finally, we can derive a unique mixed strategy equilibrium from the first case equality, with the symmetrical distribution

$$q^* = \frac{2\,(c-t)}{2\,(c-t)-C\,(2)} \quad \text{and} \quad 1-q^* = \frac{C\,(2)}{C\,(2)+2\,(t-c)}. \tag{D.2}$$

More general conditions to caracterize Nash equilibria such that both agents do mix strategies are easy to derive in this simple game, and show that the distribution exhibited in the symmetrical case actually exhausts the mixed strategy equilibria. The corresponding aggregate payoff is

$$\begin{aligned} \sum_{s\in S}\left(\prod_i p_i(s_i)\,(u_1(s)+u_2(s))\right) &= (2\,(c-t)-C(2))\,q^2 + (2q-1)\,2t - 2qc \\ &= -\frac{2t}{2t+C\,(2)-2c}C\,(2)\,. \end{aligned} \tag{D.3}$$

From our assumption on costs, we know that $C(2) < 2c$, which implies that $(q^*, 1-q^*)$ yields a smaller payoff than the socially optimal pure allocation. For a minimum tax $t = c$, the payoff under $(q^*, 1-q^*)$ is $-2c < -c$, and then strictly increases in the threat stringency with

$$\lim_{t \to +\infty} -\frac{2t}{2t + C(2) - 2c} C(2) = -C(2) \tag{D.4}$$

Specifically, mixed strategies are Pareto improving compared to the pure Nash equilibria allocations for a tax level

$$t > \frac{c(C(2) - 2c}{2(C(2) - c)} \tag{D.5}$$

Thus, extending the set of pure strategies to mixed strategies allows to reach higher expected aggregate payoffs for a high enough tax threat, provided agents find a way to coordinate on equilibria multiplicity. Let us see what would be the set of reachable payoffs in the voluntary agreement with mediated communication such as described in our coordination device. We already know that any mixed strategies Nash equilibria of a game is also a correlated equilibria of this game (Myerson 10), meaning that the VA with the coordination device certainly allows to implement $(q^*, 1-q^*)$. But we want to check if even higher payoffs could be implemented as correlated equilibria for a *given* threat level, since we know that a credibility requirement would actually limit the regulator in his choice of t. Accordingly to the general case, we denote p_{kl} the probability assigned by the regulator to the pure strategy profile $(s_1 = k, s_2 = l) \in S$, with $\sum_{s \in S} p_s = 1$. Then, following Myerson 10's interim definition of correlated equilibrium, let us write the strategic incentive constraints in our two-player game

$$\begin{aligned}
\frac{p_{11}}{p_{11}+p_{10}}\frac{C(2)}{2} + \frac{p_{10}}{p_{11}+p_{10}}c &\leq \frac{p_{10}}{p_{11}+p_{10}}t \\
\frac{p_{00}}{p_{01}+p_{00}}t &\leq \frac{p_{01}}{p_{01}+p_{00}}\frac{C(2)}{2} + \frac{p_{00}}{p_{01}+p_{00}}c \\
\frac{p_{11}}{p_{11}+p_{01}}\frac{C(2)}{2} + \frac{p_{01}}{p_{11}+p_{01}}c &\leq \frac{p_{01}}{p_{11}+p_{01}}t \\
\frac{p_{00}}{p_{10}+p_{00}}t &\leq \frac{p_{10}}{p_{10}+p_{00}}\frac{C(2)}{2} + \frac{p_{00}}{p_{10}+p_{00}}c
\end{aligned}$$

with the probability constraint

$$\begin{cases} p_{11} + p_{10} + p_{01} + p_{00} = 1 \\ p_{11} \geq 0,\ p_{10} \geq 0,\ p_{01} \geq 0 \text{ and } p_{00} \geq 0 \end{cases}, \tag{D.7}$$

which is the algebraic form for the condition that for any individual suggestion from the regulator to an agent, and provided a given probability distribution on S that was prealably announced, the agent has no incentive not to follow the suggestion. The incentive constraints rewrite

$$2p_{10}\left(\frac{t}{C(2)} - \frac{c}{C(2)}\right) - p_{11} \geq 0 \tag{D.8a}$$

$$2p_{00}\left(\frac{c}{C(2)} - \frac{t}{C(2)}\right) + p_{01} \geq 0 \tag{D.8b}$$

$$2p_{01}\left(\frac{t}{C(2)} - \frac{c}{C(2)}\right) - p_{11} \geq 0 \tag{D.8c}$$

$$2p_{00}\left(\frac{c}{C(2)} - \frac{t}{C(2)}\right) + p_{10} \geq 0 \tag{D.8d}$$

and a maximization program for the regulator can be formulated as follows

$$\max_{p_{11},p_{10},p_{01},p_{00}} -c\left(p_{10}+p_{01}\right)-C\left(2\right)p_{11}-2p_{00}t \tag{D.9}$$
$$s.t. \qquad (D.8a), (D.8b), (D.8c), (D.8d) \text{ and } (D.7).$$

Using the assumption $t > c$, and denoting correlated strategies as vectors

$$\left(p_{11}\ p_{10}\ p_{01}\ p_{00}\right)^T,$$

we first notice that the set of vectors solving program D.7 must be a subset of $\left(p_{11}\ p_{10}\ p_{01}\ 0\right)^T \in \mathbb{R}^4$. Specificaly, by substituting $p_{00} = 0$ into (D.8b), (D.8d) and (D.7), we get the set of candidates $\left(p_{11}\ p_{10}\ p_{01}\ 0\right)^T \in \mathbb{R}^4$ such that

$$\begin{cases} p_{10} \geq p_{11}\frac{C(2)}{2(t-c)} \\ p_{01} \geq p_{11}\frac{C(2)}{2(t-c)} \\ p_{11} \geq 0 \\ p_{11}+p_{10}+p_{01}=1 \end{cases} \tag{D.10}$$

Geometricaly, it is the area bounded by the inequalities

$$p_{10} \geq \frac{C(2)}{2\left(t-c\right)+C(2)} - \left(\frac{C(2)}{2\left(t-c\right)+C(2)}\right)p_{01} \tag{D.11}$$

$$p_{10} \geq 1-\left(1+\frac{2\left(t-c\right)}{C(2)}\right)p_{01} \tag{D.12}$$

$$p_{10} \leq 1-p_{01} \tag{D.13}$$

on the affine hyperplane defined by $p_{11} = (1-p_{10}-p_{01})$, with the two first inequalities intersecting in

$$p_{01} = p_{10} = \frac{1}{2}\frac{C(2)}{C\left(2\right)+t-c}, \tag{D.14}$$

as illustrated in figure D.4. Provided our assumption on costs, it is now obvious the regulator maximizes the objective by assigning a maximal probability to the full-participation profile. Consequently, the solution of the program is

$$\begin{pmatrix} p_{11}^* \\ p_{10}^* \\ p_{01}^* \\ p_{00}^* \end{pmatrix} = \begin{pmatrix} \frac{t-c}{C(2)+t-c} \\ \frac{1}{2}\frac{C(2)}{C(2)+t-c} \\ \frac{1}{2}\frac{C(2)}{C(2)+t-c} \\ 0 \end{pmatrix}, \tag{D.15}$$

and the value of the objective (or expected aggregate gain) is

$$\begin{aligned} SW_{EC}^*(t) &= \sum_{s\in S} p_s^*\left(u_1\left(s\right)+u_2\left(s\right)\right) \\ &= -c\left(p_{10}+p_{01}\right)-C\left(2\right)p_{11} \\ &= -\frac{C\left(2\right)t}{C\left(2\right)+t-c}. \end{aligned} \tag{D.16}$$

Note that both $p_{01}^* = p_{10}^*$ are decreasing in t, while $p_{11}^* = 1-(p_{01}^*+p_{10}^*)$ is increasing in t. In other words, a higher threat rises the probability on the socially optimal allocation

$$\frac{\partial}{\partial t}\left(\frac{t-c}{C\left(2\right)+t-c}\right) = \frac{C\left(2\right)}{\left(C\left(2\right)+t-c\right)^2} > 0. \tag{D.17}$$

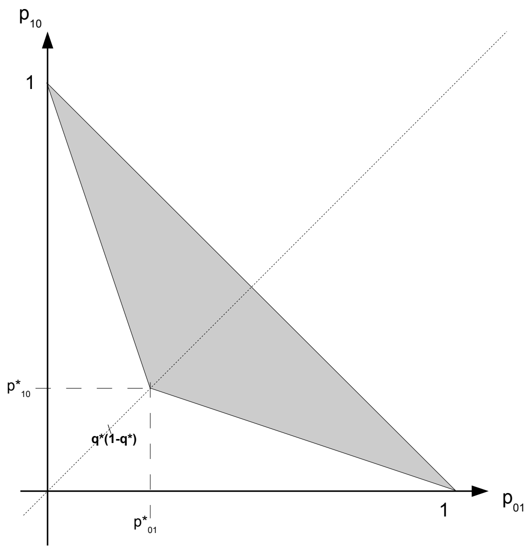

Figure D.4: The set of correlated equilibria in the two-player example.

Or, geometrically, the grey area extends along the bisectrice on figure D.4. The aggregate payoff tends then to the payoff corresponding to the socially optimal allocation

$$\lim_{t\to+\infty} SW^{*}_{EC}(t) = -C\,(2)\,. \tag{D.18}$$

Finally, let us remark as previously mentioned, that the pure and mixed Nash equilibria of the game do satisfy the correlated equilibria conditions (easily verified by substituting $p_{11} = (q^{*})^{2}$, $p_{10} = p_{01} = q^{*}\,(1-q^{*})$ and $p_{00} = (1-q^{*})^{2}$ into (D.8a)-(D.8d)). But we know now that they yield a smaller aggregate payoff than the optimal correlated equilibrium for all t. Specifically, we have the following ranking

$$SW^{*}_{EC}(t) > -2\frac{C\,(2)\,t}{C\,(2)+2\,(t-c)} > -c, \tag{D.19}$$

implying that the coordination device not only solve the problem raised by multiplicity, but also ensures that a higher expected aggregate payoff is reached for a given credible level of threat.

[1] Arora, S. and Gangopadhyay, S. (1995). Toward a theoretical model of voluntary overcompliance. *Journal of Economic Behavior and Organization*, 28(3):289–309. 2

[2] Aumann, R. J. (1974). Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1(1):67–96. 2, 5

[3] Borkey, P., Glachant, M., and Lévêque, F. (1998). Voluntary approaches for environmental policy in oecd countries : an assessment. Technical report, CERNA. 2

[4] Brau, R. and Carraro, C. (2011). The design of voluntary agreements in oligopolistic markets. *Journal of Regulatory Economics*, 39(2):111 – 142. 18

[5] Cavaliere, A. (2000). Overcompliance and voluntary agreements. *Environmental and Resource Economics*, 17(2):195–202. 2

[6] Cavaliere, A. (2001). Coordination and the provision of discrete public goods by correlated equilibria. *Journal of public economic theroy*. 3

[7] Chiambretto, A.-S. and Stahn, H. (2011). Voluntary agreements with industries - participation incentives with industry-wide targets: a comment. *Economics Bulletin*, 31(1):116–121. 13

[8] Denicolò, V. (2008). A signaling model of environmental overcompliance. *Journal of Economic Behavior & Organization*, 68(1):293–303. 2

[9] Lyon, T. P. and Maxwell, J. W. (2008). Corporate social responsibility and the environment: A theoretical perspective. *Review of Environmental Economics and Policy*, 2(2):240–260. 2

[10] Myerson, R. B. (1997). *Game Theory Analysis of Conflict*, chapter Games with communication. Harvard University Press. 9, 23

[11] OECD (1999). Voluntary approaches for environmental policy : An assessment. Technical report, OECD, Paris. 2

[12] OECD (2003). Voluntary approaches for environmental policy voluntary approaches for environmental policy: Effectiveness, efficiency and usage in policy mixes. Technical report, OECD, Paris. 2

[13] Palfrey, T. R. and Rosenthal, H. (1984). Participation and the provision of discrete public goods: a strategic analysis. *Journal of Public Economics*, 24(2):171–193. 2, 3, 7, 18, 19

[14] Segerson, K. (2013). When is reliance on voluntary approaches in agriculture likely to be effective? *Applied Economic Perspectives and Policy*, 35(4):565 – 592. 5

[15] Suter, J. F., Segerson, K., Vossler, C. A., and Poe, G. L. (2010). Voluntary-threat approaches to reduce ambient water pollution. *American Journal of Agricultural Economics*, 92(4):1195–1213. 13