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 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 shown that any symmetrical mixed Nash equilibrium (NE) of the VA without a correlation device can be implemented by 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.

[☆]The author thanks Francois Salanié and Bernard Sinclair-Desgagné for their helpful comments, as well as Hubert Stahn for the several productive discussions that helped in the starting of the present work.

[*]Corresponding author

Email address: annesarah.chiambretto@univ-amu.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

1. Introduction

Voluntary agreements (VAs hereafter) generically refer to the many and multiform schemes whereby agents that generate some environmental externality engage in self-regulation. When these proactive behavior are motivated by a background regulatory threat, whether enacted (e.g. when used in a policy mix coupled with command and control, see Borkey et al. 3) or merely potential (Glachant 9), the literature mention them as *preemptive*. Since potential enactments most often turn out to be sectoral, preemptive VAs implicitly rely on a collective liability rule. Namely, if some global environmental cap cannot be voluntarily achieved by a group, the mandatory legislation will apply regardless of individual voluntary efforts. Such voluntary regulation mechanisms have been widely studied in game-theoretic frameworks with no inter-firms communication (e.g. see Segerson and Wu 14, Dawson and Segerson 8, or Brau and Carraro 4).

The present work deals with preemptive VAs, focusing on firms' participation decision and its properties when the VA comes along with what we call 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 VA. These gains depend on the regulatory threat t, and the cost of providing the public good (i.e. the cost of satisfying the 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. Specifically, if agents' strategy sets correspond to the sets of options naturally present in the non-mediated game, and agents coordinate on an exogenous random signal, then our results can be seen as a description of inter-firm communication in self-regulation initiatives. If conversely, agents' strategy sets are seen as a menu defined by (or negociated with) the regulatory agency, then our results give rise to normative implications. Our base game may thus fit the three main categories of VAs identified by OECD's classification[2] (see OECD 12). For all the institutional forms aforementioned we find that, if instead of offering participation to all the firms from the sector to be regulated, a third party makes private recommendation to each firm, it can implement a better participation equilibrium than the one that could have been possibly achieved in a case without private recommandation. We characterize such an equilibrium in the general N-player case.

. 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.

[2]A widely adopted classification distinguishes between three main categories[3], set on the basis of the regulatory agency involvement level (OECD 12) : (i) *public voluntary 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 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 VAs with an implicit collective liability. Otherwise, Cavaliere 5 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. Last but not least, Arce and Sandler 1 make a similar point by relating CE to International Environmental Agreements (IEAs). While undoubtedly fruitful, their contribution consists of highlighting conceptual correspondances and illustrating it by rudimentary 3-players examples under different aggregation technologies. Moreover, while central to the present paper, the specific issue of the threat is not investigated since irrelevant in the application context of IEAs.

. This work is organized as follows. In section 2, we present the non mediated VA and the correlation device. Section 3 elaborates on some crucial hypothesis regarding strategies implemented by the correlation device. The socially optimal allocation and the Nash equilibria of the non-mediated VA are characterized in section 4 as benchmarks. In section 5, we study the correlated equilibria and identify the optimum. Section 6 relates the unique mixed strategy Nash equilibrium to the set of correlated equilibria, before we provide a full study of the impact of the threat on welfare gains generated by the VA. Finally, we numerically illustrate our results in a specified example of pollution abatement in section 7.

2. The model

2.1. The basic setting

Consider a regulator willing to achieve a global regulation cap that involves the production activities[4] of N identical firms, indexed by $i \equiv \{0, ..., N\}$. We then assume some third party offers each firm to participate to the target implementation by subscripting to a VA. Firm i can either accept ($s_i = 1$) or reject ($s_i = 0$) the offer, knowing that participation modalities depend on the total number of participating firms, $\sum_i s_i \equiv m$.

[4]Such a cap may be for instance to conform some given share of a sector output with an efficiency standard (e.g. the ACEA agreement), or to reduce targeted agents' toxically releases to some limit value (e.g. the EPA's 33/50 Program).

Specifically, for any fixed cap level, the total cost of its implementation is given by a function $C : m \to \mathbb{R}^+$ decreasing in the number of participating firms:

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

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

$$(m\ c_i\,(m))' = m\ c_i'\,(m) + c_i\,(m) < 0. \tag{2}$$

As for the 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[5].

The regulator may consider the achievement of a given target by too few agents is cost-inefficient or not feasible. Thus, if a participation threshold $w \in \mathbb{N}^+$ is not reached, the VA fails and a collective tax[6] is enforced. Conversely, if the participation threshold is achieved, the VA succeeds and the $m \geq w$ participating firms implement the global target, equally sharing the total implementation costs, $C(m)$. Therefore, agent i's payoffs are defined by

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

where $t > t_w := {}^{C(w)}/_{w}$ is the individual payoff under the mandatory tax. Such an assumption on t actually ensures that w satisfies a profitability condition in the sense of self-enforcing equilibrium[7] and, as a direct consequence here, that the Nash solution is non trivial. Before we present the correlation device, let us stress that w and t are considered as exogenous, so as to keep the analysis as general as possible. Thus, the so called third party may as well be embodied by the regulator in the case of a public VA, or an industry association in the cases of self-regulation and negotiated VAs. For instance, consider the third party as an industry association which centralizes and coordinate firms' self-regulation incentives. The participation threshold would then be constrained to ${}^{C(w)}/_{w} =$

[5] Let $c_i(\rho)$ denote the total implementation cost, by agent i, of some fixed emission reduction burden, ρ. If $c_i(\rho)$ is increasing and convex in ρ, then ${}^{c_i(\rho)}/_{\rho}$ is increasing in ρ (decreasing in the corresponding cap). It follows $\rho \times {}^{c_i(\rho/m)}/_{\rho/m}$ is also decreasing in m, which we can rewrite, for a fixed cap level, $c_i\,(m)\,m$.

[6] By collective, we mean it applies to each of the N agents whatever their individual willingness to undertake voluntary action may be.

[7] A coalition is said to be self-enforcing (d'Aspremont et al. 7, Dawson and Segerson 8) iif participation is (i) profitable, (ii) internally and externally stable. While (ii) is conceptually equivalent to a Nash equilibrium condition, profitability tests for simultaneous (as opposed to unilateral) participating agents' deviations. In public good games, the use of a breaking rule conjointly with a assumption on payoffs such that it is better to participate than not to participate at the threshold (here $t > C(w)$), allow to mimic coalitional incentives within a non-cooperative framework.

t, i.e. the minimum profitable level of participants for some given level of t. Such a refinement remains a subcase of our general game.

2.2. The correlation device

. Now, suppose that instead of offering voluntary participation to each agent, the third party privately recommends either $s_i = 0$ or $s_i = 1$, depending on index i. Let us denote by $S = \times_{i \in N} S_i$ the set of strategy profiles, where $S_i = \{1, 0\}$ is the individual strategy set. The recommendation is the result of a random selection process on S, the distribution of which, a mapping $p : S \to [0, 1]$ with $\sum_{s \in S} p(s) = 1$, is assumed to be public knowledge. Suppose for example, that distribution p is prealably announced by the third party, while each agent is only told his individual strategy s_i in the selected profile $(s_i)_i \equiv s$. Again, she can freely follow or reject the private recommendation, and the VA succeeds as long as at least w agents participate. Such a regulatory device is a straightforward application of correlated strategies under private[8] signaling, as developed in Aumann 2.

Finally, let us denote by S_k the set of strategy profiles such that the total number of participating agents is k, ie. formally: $S_k = \{s : \sum_{i=1}^{N} s_i = k\}$. We decide to restrict our analysis to $\psi(S)$, defined as the subset of distributions (or correlated strategies[9]) such that participation profiles with same number of participants are equiprobable. To be more specific:

Definition 1. *In this paper, the terms of 'mediated' or 'transformed' game will interchangeably refer to the VA featuring correlated strategies which satisfy, unless mentioned otherwise, an extra hypothesis of equiprobability. Namely, the correlation device is assumed to implement $p \in \psi(S)$, where $\psi(S)$ is the set composed only of distributions that verify:*

$$p(s) = {}^{1}/_{\binom{k}{N}} \sum_{s \in S_k} p(s), \qquad \forall s \in S_k, \quad k \in \{0, ..., N\}. \tag{4}$$

The previous definition will avoid having us consistently to mention the results hold under the equiprobability assumption. Next section is devoted on elaborating on condition (4), which fully characterizes our restriction (as opposed to Aumann 2's general definition which holds for any probability distribution).

3. Preliminary results and definitions

Here, we detail some crucial consequences of the equiprobability assumption on participation rates and the payoffs structure. To do so, let us remark that $\cap_{k \in \{0,...,N\}} S_k$ is an exhaustive set of events in the probability space $(S, \mathcal{P}(S))$ defined by the correlation

[8] Some of the forthcoming preliminary results may be also reachable through a public signaling device though.

[9] One should know that *correlated strategy* is simply the term which stands for *distribution* in the framework of mediated games, both designations may thus be used interchangeably in some parts of the present work.

device, where $\mathcal{P}(S)$ is the power set of S. Therefore, the number of participating agents m can be defined as a random variable, and its distribution inferred from p as follows:

$$pr(m = k) = \sum_{s \in S_k} p(s), \quad k \in \{0, ..., N\}. \tag{5}$$

It follows the set S_k can be partitioned according to the value of s_i, with:

$$pr\left(m = k\right) = pr\left(m = k \cap s_i = 0\right) + pr\left(m = k \cap s_i = 1\right) \tag{6}$$

for all $k \in \{0, ..., N\}$. Note, besides, that provided some p on the set of strategy profiles S, one can easily derive the distributions of individual strategies s_i as the marginals of a joint probability distribution. Specifically, for all i, marginal distributions are given by $pr(s_i = 1) = \sum_{s_{-i} \in S_{-i}} p(s_i = 1, s_{-i})$ and $pr(s_i = 0) = \sum_{s_{-i} \in S_{-i}} p(s_i = 0, s_{-i})$ where $\cup_{s_i \in S_i} \left(s_i \cap s_{-i}\right)_{s_{-i} \in S_{-i}}$ is also a complete system of events in $(S, \mathcal{P}(S))$. We can now close this discussion with a simple observation. One interesting implication of the equiprobability assumption, is that the corresponding marginal distributions are symmetric. Especially, we state the following useful result.

Lemma 1. *For all $p \in \psi(S)$, the individual probability to participate is given by: $pr(s_i = 1) = \sum_{k=0}^{N} \frac{k}{N} pr(m = k)$ for all i, where k/N is the probability that agent i is told by the regulator to play $s_i = 1$, provided the randomly selected profile features k participants.*

Proof 1. Using the partition of S_k defined in eq. (6), let us rewrite the marginal distribution of i's participation: $pr(s_i = 1) = \sum_k p(m = k \cap s_i = 1)$. Then notice that some elementary combinatorics yield:

$$pr(s_i = 1 \,|\, m = k) = \tbinom{k-1}{N-1} / \tbinom{k}{N} = \frac{k}{N}, \tag{7}$$

which we can, from the conditional probabilities definition, substitute into the new expression for $pr(s_i = 1)$. □

Note, however, that condition (4) is sufficient yet not necessary for the marginal distributions to be symmetric[10]. From this lemma, useful results about the individual payoff and the expected number of participating agents can be derived.

3.1. The conditional payoffs

After a given distribution $p \in \psi\left(S\right)$ is announced, agent i's *expected* utility is given by: $\mathcal{U}_i^p = \sum_{s \in S} p(s) u_i(s_i, m)$. Then, using the conditional distribution of s_i, given a number of participants $m = k$, defined by $\sum_{s_i \in \{0,1\}} pr(s_i \mid m = k) = 1$, we obtain:

$$\mathcal{U}_i^p\left(t\right) = -\sum_{k=0}^{w-1} pr(m = k)t - \sum_{k=w}^{N} pr(m = k)\frac{C(k)}{N}. \tag{8}$$

[10]For this reason, we deliberately do not use **?** 's fallacious terminology of 'symmetrical' correlated strategies/equilibrium etc.

In other words, agent i's expected utility is a continuous functional of t, the level of threat. Besides, since each player may also be able to revise the distribution on S *conditionally* to the prescription that was privately made to him by the third party, we denote by $\mathcal{U}_i(p \mid s_i)$ his expected utility contingent upon s_i. Specifically, using lemma 1, we obtain for $s_i = 1$:

$$\mathcal{U}_i(p \mid s_i = 1) = -\sum_{k=0}^{w-1} \frac{k}{N} \frac{pr(m=k)}{pr(s_i=1)} t - \sum_{k=w}^{N} \frac{pr(m=k)}{pr(s_i=1)} \frac{C(k)}{N} \tag{9}$$

since $pr(m = k \,|\, s_i = 1) = {}^{k}/_{N} \frac{pr(m=k)}{pr(s_i=1)}$.

3.2. The expected number of participating agents

Before we turn to the equilibria analysis, let us mention two last results of interest, that may also convey valuable information from the regulator perspective.

Lemma 2. *For all $p \in \psi(S)$, the expected number of participating agents to the VA is given by: $E(m) = Npr(s_i = 1)$.*

Proof 2. Observe that: $E(m) = \sum_{k=0}^{N} k\ pr(m = k) = N \sum_{k=0}^{N} \frac{k}{N} pr(m = k)$, where the sum is the marginal distribution of $s_i = 1$ defined in lemma 1.

The equiprobability assumption implies that the expected number of participating agents linearly depends on the individual probability to participate. Besides, the expected aggregate payoff simply rewrites:

$$\sum_i \mathcal{U}_i^p(t) = N\mathcal{U}_i^p(t) := SW_p(t).$$

Indeed, by construction of the third party's device, symmetric agents face, ex-ante, the same distribution. Associated with lemma 2, the expression of the aggregate welfare paves the road for the final efficiency analysis, allowing us to establish a univocal relationship, for all t, between participation rates and $SW_p(t)$, under our minimal costs assumptions.

. We now turn to the equilibria analysis. In particular, next sections' results will be used as a basis for comparison to the case of the game fitted with the correlation device.

4. The outcome of the non-mediated VA

Studying the outcome of the base game amounts to characterize the pure and mixed Nash equilibria. Let us provide a full analysis.

4.1. Multiplicity of pure Nash equilibria

. We first focus on the pure Nash equilibria (PNE) of the game. Beforehand, regarding degenerate cases of distribution p such that $p(s) = 1$ and $p(s') = 0$ for all $s' \neq s$, let us point out the following:

Remark 1. *The full participation profile,* $(1)_{i=1}^{N}$*, yields the highest aggregate payoffs, and more generally, for all* $s \neq s'$ *in* S*, we have:* $m' \geq m$ *if and only if* $u_i(s') \geq u_i(s)$*, where* $\sum_i s'_i = m'$ *and* $\sum_i s_i = m$*.*

This statement directly follows from the costs assumptions. As a result, deterministic strategy profiles can be Pareto-ranked according to the number of participating agents.

We now 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 $S_{-i} \equiv \times_{j \neq i} S_j$ with $s_{-i} \in S_{-i}$, the set of feasible opponents' pure strategy profiles. For all agent i:

$$\Phi_i(s_{-i}) = \begin{cases} \{0,1\} & \text{if } \sum_{j \neq i} s_j < w-1 \\ 1 & \text{if } \sum_{j \neq i} s_j = w-1 \\ 0 & \text{if } \sum_{j \neq i} s_j > w-1 \end{cases}. \tag{10}$$

Note that $\Phi_i = argmax_{s_i \in \{0,1\}} u_i(s_{-i}, s_i)$ can be redefined as a correspondence from $\mathbb{N}^+$ to $\mathcal{P}(S_i)$, since only the number of participating firms matters in the payoffs definition, and let us denote S^{NE} the set of PNE, namely strategy profiles s^{NE}, defined by $\left(\Phi_i(s^{NE}_{-i})\right)_i = s^{NE}$.

Remark 2. *If* $N \geq 2$ *and* $w \geq 2$*, the* $|S^{NE}| = \binom{w}{N} + \sum_{k=0}^{w-2} \binom{k}{N}$ *PNE of the public VA without the correlation device are given by* $S^{NE} = \{S_w, (S_k)_0^{w-2}\}$*, none of which corresponds to the socially optimal allocation,* $(1)_{i=1}^{N}$*. Otherwise,* $|S^{NE}| = \binom{w}{N}$ *and* $S^{NE} = \{S_w\}$*.*

Proof 3. Observe that any strategy profile such that $m < w-1$ *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* $m = w-1$*, each non-participating agent has a unilateral incentive to deviate and pay* $C(w)/w$ *instead of the tax* t*. Likewise, for any profile such that* $m > w$*, 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* $m = w$ *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.*

Even though we identified PNE such that the VA succeeds, multiplicity raises the question of the public VA feasibility : how will agents coordinate amongst the several subsets of Pareto equivalent PNE ? This issue will be adressed in section 6.

4.2. The unique symmetric MNE

. Let us conclude our benchmark study with an analysis of the symmetric mixed Nash equilibria (MNE) of the base game. Specifically, consider that the N agents of the VA are mixing in the support $\{0,1\}$ according to a symmetric marginal distribution, $p(s_i = 1) = q$ and $p(s_i = 0) = 1 - q$, for all i. By definition of MNE, individual strategies are assumed to be played independently. It follows the corresponding joint distribution on the set of participation profiles, S, is given by:

$$p(s_k) = q^k (1-q)^{N-k} \quad \forall s_k \in S_k, \qquad k \in \{0, ..., N\}.$$

As the base game is finite and symmetric, we know there alway exists[11] a symmetric MNE, and that it is the mixed strategy profile such that the individual participation probability, q, satisfies:

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

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[12], i.e. which also includes partial supports, is provided in Appendix A. From condition (11), we establish the following proposition:

Proposition 1. *The VA without the correlation device has a unique symmetric MNE, given by* $\{Q(t) := pr(s_i = 1), (1 - Q(t)) := pr(s_i = 0)\}$*, with* $Q'(t) > 0$*,* $Q(0) = 0$*, and* $\lim_{t \to +\infty} Q(t) = 1$*, where* $Q(t)$ *is defined as the inverse function of:*

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

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

$$\frac{\partial}{\partial q} t(q) = \sum_{k=w-1}^{N-1} \frac{\binom{k}{N-1}}{\binom{w-1}{N-1}} \frac{C(1+k)}{1+k} \left[\frac{1}{(1-q)^2}\right]^{k-w+1} > 0. \tag{13}$$

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

Now that we have characterized the MNE participation probability, $Q(t)$, and its relationship with the threat t, let us focus on the participation rate of the non-mediated VA. In particular, the following can be stated:

[11] See **?** for a complete proof of the existence of symmetric NE in finite and symmetric games.

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

Corollary 1. *The joint distribution, $p^N \in \psi(S)$, corresponding to the marginal MNE distributions $\{Q(t), (1-Q(t))\}$, is defined by: $p^N(s_k) = Q(t)^k(1-Q(t))^{N-k}$ for all $s_k \in S_k$, with $k \in \{0, ..., N\}$. As a result, the expected number of participating agents, $E(t, N) = Q(t)N$, strictly increases in $t > 0$.*

This last result straightforwardly follows from Lemma 2 and Proposition 1.

We now turn to the studying of the outcome of the game fitted with the correlation device. The relevant equilibrium concept is that of correlated equilibrium (CE).

5. Characterizing the optimal correlated equilibrium

In the present section, we first define the set of CE of the mediated VA game, i.e. distributions $p \in \psi(S)$ which satisfy both the probability and the incentive constraints such as defined by Myerson 10. We then identify a CE in $\psi(S)$ that generates the highest aggregate payoffs.

5.1. The set of CE

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

$$-\sum_{k=0}^{k\leq w-1} pr\left(m = k \mid s_i = 0\right) t - \sum_{k=w}^{N-1} pr\left(m = k \mid s_i = 0\right) 0 \geq -\sum_{k=0}^{k\leq w-2} pr\left(m = k \mid s_i = 0\right) t - \sum_{k=w-1}^{N-1} pr\left(m = k \mid s_i = 0\right) \frac{C(k+1)}{k+1}$$

must holds for all i, which simplifies to:

$$pr\left(m = w-1 \mid s_i = 0\right)\left(t - \frac{c}{w}\right) - \sum_{k=w}^{N-1} pr\left(m = k \mid s_i = 0\right) \frac{C(k+1)}{k+1} \leq 0. \qquad (14)$$

Such a condition ensures agent i's expected payoff is higher if she complies than if she decides to participate, provided her prior (i.e. the announced distribution), has been revised taking into account the fact that the profile the third party selected necessarily features $s_i = 0$. 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:

$$pr\left(m=w \mid s_i=1\right)\left(t-\frac{c}{w}\right)-\sum_{k=w+1}^{N} pr\left(m=k \mid s_i=1\right)\frac{C(k)}{k}\geq 0. \tag{15}$$

Again, condition (15) guarantees agent i's expected payoff is higher if she complies than if she decides not to participate, provided her prior has been revised taking into account that the profile the regulator selected necessarily features $s_i=1$. Finally, the probability constraints are given by:

$$p\in\psi\left(S\right)\Leftrightarrow\begin{cases}\sum_{k=0}^{N} pr\left(m=k\right)=1\\ \left(p_{(0)_N},...,p_{(1)_N}\right)\in\mathbb{R}^{N+}\\ \text{and } p(s)=\frac{pr(m=k)}{\binom{k}{N}}, \quad \forall s\in S_k, \qquad k\in[0,N],\end{cases} \tag{16}$$

where the last equality stands for the fact our VA features equiprobable correlation devices solely[13]. Using lemma 1, we know such a restriction on p actually amounts to substitute

$$pr(m=k\,|\,s_i=1)=\frac{k}{N}\frac{pr\left(m=k\right)}{pr\left(s_i=1\right)}$$

into the incentive constraints (14) and (15), which therefore conveniently rewrite:

$$\begin{cases} pr\left(m=w-1\right)\left(N-w+1\right)\left(\frac{C(w)}{w}-t\right)+\sum_{k=w}^{N-1}(N-k)\ pr(m=k)\frac{C(k+1)}{k+1}\geq 0\\ w\ pr\left(m=w\right)\left(t-\frac{C(w)}{w}\right)-\sum_{k=w+1}^{N} pr\left(m=k\right)C\left(k\right)\geq 0.\end{cases}$$

These two last inequalities, conjointly with the probability constraints, fully characterize the set of CE of the mediated VA.

5.2. The optimal CE

Now consider the third party seeks to implement some distribution p on the set of strategy profiles which maximizes:

$$\max_{p\in\psi(S)} -Nt\sum_{k=0}^{w-1} pr\left(m=k\right)-pr\left(m=w\right)C(w)-\sum_{k=w+1}^{N} pr\left(m=k\right)C(k) \tag{17}$$

Such a distribution should be incentive compatible to be workable, i.e. it has to verify the conditions (14) and (15) characterized above. The third party therefore maximizes

[13]The set of constraints without condition (4) characterizes the general set of CE of the participation game.

the objective (17) under the constraints:

$$\text{s.t}\begin{cases} pr(m=w-1)\,(N-w+1)\left(\dfrac{C(w)}{w}-t\right)+\displaystyle\sum_{k=w}^{N-1}(N-k)\;pr(m=k)\dfrac{C(k+1)}{k+1}\geq 0 & \text{(18a)}\\ w\;pr\,(m=w)\left(t-\dfrac{C(w)}{w}\right)-\displaystyle\sum_{k=w+1}^{N} pr\,(m=k)\,C\,(k) \geq 0 & \text{(18b)}\\ \text{and}\quad \displaystyle\sum_{k=0}^{N} pr\,(m=k) =1. \end{cases}$$

We thus recognize a linear programming problem, which can be converted into its augmented form by introducing two slack variables, denoted x_1 (into constraint (18a)), and x_2 (into constraint (18b)). Note that the generated standard program is only composed of $j = 3$ equality constraints and $l = N + 2$ variables, since the symmetry hypothesis implies the objective is actually optimized in $\{pr(m = k)\}_0^N$. Such a program can be solved by applying the two-phase simplex algorithm**?** . However, this method may imply numerous (and cumbersome) iterations if 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 B) argument.

Proposition 2. *The optimal CE of the mediated VA, denoted p^*, is the probability distribution on the set of participation profiles, S, given by:*

$$p^*(s) = \begin{cases} \frac{(tw-C(w))}{(tw-C(w))+C(N)} & \text{for } \; s=(1)_{i=1}^{N}\,, \\ 1/\binom{w}{N}\frac{C(N)}{(tw-C(w))+C(N)} & \text{for all } \; s\in S_w, \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

Proof 5. First, observe that (18a) and (18b) do not depend on $\{pr(m = k)\}_0^{w-2}$. Since $t > c$, it is straightforward the optimal distribution must assign a probability 0 to the corresponding participation profiles. Then, remark that $pr(m = w-1)$ appears in (18a) solely, and that it does not need to be strictly positive. However, as $pr\,(m = w)$ must be strictly positive for (18b) to hold, we know (18a) cannot be equal to 0 at the optimum. Finally, $\frac{\partial C(N_p)}{\partial N_p} < 0$ implies (18b) will be binding, since it is now obvious the highest probability should be put on profile $(1)_i$. We therefore know $\{x_1,\ pr(m=w),\ pr(m = N)\}$ is the basis which has to be tested in order to confirm our heuristic solution is optimal. This is done in Appendix B. Finally, eq. (7) of Lemma 1 is used to build the corresponding joint probability distribution on S.

Moreover, we know from Lemma 1, that the corresponding optimal marginal distributions are given by:

$$pr^*(s_i = 0) = \frac{N-w}{N}\frac{C(N)}{(tw-C(w))+C(N)} \tag{20a}$$

$$pr^*(s_i = 1) = \frac{w}{N}\frac{C(N)}{(tw-C(w))+C(N)} + \frac{(tw-C(w))}{(tw-C(w))+C(N)}, \tag{20b}$$

where $pr^*(s_i = 1)$ can be reinterpreted as $\frac{E^*(t,N)}{N}$, the expected participation rate under the optimized correlation device.

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

6. Efficiency analysis

6.1. The coordination issue

Consider the unique marginal distribution $\{Q(t), (1 - Q(t))\}$ such that the condition for symmetric MNE holds. Then, it is known that the corresponding joint distribution p^N also verifies conditions (14) and (15), i.e. the general CE strategic incentive constraints (Aumann 2, Nisan et al. 11). In particular, since the equiprobability condition (4) also holds, we show p^N actually belongs to the subset $\psi(S)$, within which it *uniquely* satisfies an extra-assumption of individual participation decisions independence. A first practical result about the potentialities of correlated participation to VAs is thus the following:

Proposition 3. *The symmetric MNE of the non-mediated VA can be implemented by a third party that uses the correlation device.*

Proof 6. See Appendix C.

In other words, the link established between the set of CE and the symmetric MNE addresses the issue raised in Remark 2 : adding the correlation device allows to overcome the coordination defect, that may arise when a public VA is implemented. Even more interestingly, since we know, from proposition 2, that p^* is an optimum over $\psi(S)$ for all t, it follows both distributions can now be ranked in terms of expected social welfare.

Proposition 4. *Let us denote by $SW_C^*(t)$ and $SW_N(t)$, the aggregate payoff, $SW(t)$, evaluated at the distributions p^* and p^N respectively. It can be stated:*

$$SW_C^*(t) \geq SW_N(t), \qquad \forall t, \tag{21}$$

or, equivalently, that p^* pareto-dominates p^N for all given level of threat t.

Proof 7. The distribution p^* is an optimal EC in $\psi(S)$ (see proposition 2). Added to proposition 3 above, it straighforwardly implies p^* pareto-dominates p^{NE}.

Indeed, as $p \in \psi(S)$, individual expected utilities are symmetrical, hence an increase of the collective welfare is ensured to actually constitutes a Pareto-improvement. Finally, let us notice that even though agents are able to coordinate on the symmetric MNE, the latter relates strictly positive probabilities to participation profiles such that the VA actually fails. Specifically, when the distribution implemented is p^N, the voluntary policy failure occurs with probability $\sigma(t) \equiv \sum_{k=0}^{w-1} Q(t)^k (1 - Q(t))^{N-k}$. On the contrary:

Corollary 2. *With the optimal correlation device, the VA always succeeds and the aggregate expected payoff is given by:*

$$SW_C^*(t) = \frac{-twC(N)}{tw - C(w) + C(N)}. \tag{22}$$

6.2. Impact of the threat

We pursue the efficiency analysis by investigating the impact of the threat on the pareto-ranking of the VA, with and without the optimal correlation device.

Proposition 5. *Under p^N, there exists a minimum threat level, $t_1 > t_w$, beyond which an increase of t leads to an increase of the expected aggregate payoffs, $SW_N(t)$. Moreover, let $t_2 \geq t_1$ denote the unique threat level such that:*

$$SW_N(t_2) = SW_N(t)|_{t \in \{\underset{t\in[0,t_1]}{argmax}\ SW_N(t)\}}.$$

Then, a regulatory threat t, pareto-dominates another t', if $t \geq t'$ and $t \geq t_2$ (sufficient condition).

Proof 8. Observe that: $\partial p(s_k)/\partial q = kQ^{k-1}(1-Q)^{N-k} - (N-k)(1-Q)^{N-k-1}$, 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, assuming $t > t_w$, i.e. $E(t,N) > k$, we observe that an increase of the threat induces two effects on the expected payoff generated by participation profiles such that the VA fails. On one hand, $\sigma(t)$ decreases since $t > t_w$, for the benefit of more cost efficient participation profiles, which induces an increase of $SW_N(t)$. On the other hand, the tax applied in case of failure is higher, which lowers $SW_N(t)$. We show that for some $t_1 > t_w$, the first effect dominates the second one (see Appendix D for a detailed proof). Finally, by construction of t_2, the sufficient condition directly follows from the first statement of the proposition.

. Second, let us focus of on the mediated VA. But before dealing with distribution p^*, we want to study the impact of the threat on the set of CE. To do so, let us rewrite (18a) and (18b) as follows:

$$\sum_{k=w+1}^{N} \frac{N-(k-1)}{N-(w-1)} \left(\frac{pr(m=k-1)}{pr(m=w-1)}\right) \frac{C(k)}{k} \geq t - \frac{C(w)}{w} \geq \sum_{k=w+1}^{N} \frac{k}{w} \left(\frac{pr(m=k)}{pr(m=w)}\right) \frac{C(k)}{k}, \quad (23)$$

where each ratio of probabilities can be interpreted as a deviation's pivotality rate.

Remark, then, that a more stringent tax does not impact the RHS inequality, i.e. the unilateral incentives to comply when the third party recommends $s_i = 1$. Indeed, if there were already no incentive not to participate for some t, then a firm will be even less willing to take the risk the VA fails when the threat is higher. However, an increase of t does directly impact the LHS. It can be seen as the amount the firm would gain by not participating when $s_i = 0$ is prescribed, and must therefore remain smaller than the deterrent income 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 game with the correlation device [14]

. We now turn to studying the impact of the threat on the optimal CE specifically. A first result of interest regards the expected number of participating agents.

Proposition 6. *Under the optimal mediation, an increase of the threat rises the expected participation rate.*

Proof 9. We know from lemma 2 that, when p^* is implemented, the participation rate actually corresponds to $pr^*(s_i = 1)$. Then, using the marginal distribution (20b) defined in the previous section, it can be shown that:

$$\begin{aligned}\frac{\partial}{\partial t}pr^*(s_i = 1) &= \frac{w}{N}\frac{\partial}{\partial t}pr^*(m = w) + \frac{\partial}{\partial t}pr^*(m = N) \\ &= \frac{tC(N)\left(N - w\right)}{N\left(\left(tw - C\left(w\right)\right) + C(N)\right)^2} > 0,\end{aligned}$$

i.e. the probability to participate strictly increases in t under our assumptions.

Under our minimal assumptions on costs, an increase of the expected number of participating firms does not necessarily imply a better expected aggregate payoff though. Next proposition establishes the impact of the threat stringency on $SW_C(t)$.

Proposition 7. *When the third party implements p^*, a regulatory threat t (strictly) pareto dominates another t' if and only if $t \geq (>)\ t'$. In particular, as t increases, the expected aggregate payoff tends toward the level of welfare generated by the full participation profile, $-C(N)$.*

Proof 10. Under our assumptions:

$$\frac{\partial}{\partial t}SW_C^*(t) = \frac{-wC\left(N\right) + w^2tC\left(N\right)}{\left(tw - C\left(w\right) + C\left(N\right)\right)^2} > 0, \tag{24}$$

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

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

Thus, the expected aggregate payoffs increases with the expected participation rate under the optimal correlation device as well. The last result of this section characterizes t_1:

[14] This result is consistent with Chiambretto and Stahn 6 and Suter et al. 15, in which the authors show inter alia that an increase of the unit tax level on emissions in a Counot game, may lead to under-participation to some ex-ante preemptive (non-mediated) VA.

Proposition 8. *The difference between the expected aggregate payoffs generated by p^N and by the optimal mediation decreases in $t > t_1$.*

Proof 11. Proposition 1 implies $\lim_{t\to+\infty} Q(t)^N = 1$ *and* $\lim_{t\to+\infty}(1-Q(t))^{N-k}Q(t)^k = 0$ *for all* $k < N$*, hence:*

$$\lim_{t\to+\infty} \sum_{w}^{N} Q(t)^k (1-Q(t))^{N-k} C(k) = \lim_{t\to+\infty} SW_C(t) = -C(N).$$

Moreover we know, from proposition 5 and eq. (24), that both aggregate payoffs, $SW_C^*(t)$ *and* $SW_N(t)$*, increase in* $t > t_1$*, which, considered with inequality (21), implies in turn* $\lim_{t\to+\infty} SW_N(t) - SW_C^*(t) = 0$.

Even though $SW_N(t)$ tends towards the expected aggregate payoffs under the optimized correlation device as t tends to infinity, let us stress that the threat may be constrained when applied to a specified model. This is the case in next section's example.

7. A numerical example

We apply the mediated VA game to a specified pollution abatement model in order to numerically illustrate last section's general results.

7.1. The pollution abatement model

Consider N symmetric firms producing a good and pollutant emissions which are engaged in Cournot competition. Let $\pi(e_i)$ be the indirect profit function, with $\partial^2\pi(e_i)/\partial^2 e_i \leq 0$. Specifically, assume

$$\pi(e_i) = kae_i - kbe_i^2 + a^2(1-k)/4b, \tag{25}$$

with $a, b > 0$, and $e^{LF} = a/2b$ the optimal level of emission at laisser faire. Parameter k and the constant will allow us to control for concavity (or sensitivity to emissions) in simulations without affecting the laissez-faire equilibrium. Let us denote $E \equiv N\epsilon$ the emission target, which amounts to an aggregate reduction of $N(e^{LF} - \epsilon)$ emissions units, with $\epsilon \in [0, e^{LF}]$. Assume furthermore the regulator chooses w such that it is feasible, i.e. $w = w_F$ where w_F is defined by $w_F = \lceil N(1 - \epsilon/e^{LF})\rceil$. As in the general participation game, let us denote by t the tax threat. It follows the payoffs of firm i in the basic game are given by :

$$u_i(s) = \begin{cases} -\frac{k}{b}\left(\frac{N(a-2b\epsilon)}{2m}\right)^2 & s_i = 1 \text{ and } m \geq w \\ 0 & s_i = 0 \text{ and } m \geq w \\ -t & m < w \end{cases} \tag{26}$$

where $C(m)/m = \left(k(N(a-2b\epsilon))^2/b(2m)^2\right)$ is the individual participation cost which is decreasing in m from individual profits' concavity. To be more specific:

$$-\frac{C(m)}{m} = -\left(\pi\left(e^{LF}\right) - \pi\left(\frac{N\epsilon - (N-m)e^{LF}}{m}\right)\right) = -\frac{k}{b}\left(\frac{N(a-2b\epsilon)}{2m}\right)^2.$$

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 laisser 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:

$$t^{PIG} = \left.\frac{\partial \pi\left(e_i\right)}{\partial e_i}\right|_{e_i=\epsilon} = k\left(a - 2b\epsilon\right). \tag{27}$$

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

$$\begin{aligned} t &= \pi(e^{LF}) - \pi\left(\epsilon\right) + t^{PIG}\epsilon + \tau \\ &= \frac{a^2 k}{4b} - bk\epsilon^2 + \tau. \end{aligned}$$

We assume furthermore $\tau \in [\underline{\tau}, \overline{\tau}]$, where $\overline{\tau}$ corresponds to a zero profit condition under the pigouvian tax :

$$\begin{aligned} \overline{\tau} &= \pi\left(\epsilon\right) - t^{PIG}\epsilon \\ &= \frac{a^2(1-k) + 4b^2 k\epsilon^2}{4b}. \end{aligned}$$

Lower bound $\underline{\tau}$ stands for the minimum tax level such that the profitability condition is satisfied at $w = w^F$, i.e. formally: $w_p(\underline{\tau}) = w_F$, where w_p is the minimum profitable participation level such as defined in the next lemma.

Lemma 3. *The minimum participation level such that participation is profitable, denoted $w_p(\tau)$, is implicitly characterized by*

$$\pi\left(\frac{N\epsilon - \left(N - w_p(\tau)\right)e^{LF}}{w_p(\tau)}\right) - \pi\left(\epsilon\right) + \epsilon\frac{\partial \pi\left(e_i\right)}{\partial e_i}\,|_{(e_i=\epsilon)} + \tau = 0, \tag{28}$$

and is decreasing in the tax threat, for all $\tau \geq 0$.

Proof 12. Using the theorem of implicit function, we get:

$$\frac{\partial w_p\left(\tau\right)}{\partial \tau} = -\frac{\left(w_p\left(\tau\right)\right)^2}{\frac{\partial}{\partial e_i}\pi\left(e_i\right)\,|_{\left(e^{LF} - \frac{E^{LF}-E}{w_p(\tau)}\right)}\left(Ne^{LF} - E\right)} \leq 0, \tag{29}$$

where $\frac{\partial^2 \pi}{e_i} < 0$ implies $\frac{\partial}{\partial e_i}\pi\left(e_i\right)\,|_{\left(e^{LF} - \frac{E^{LF}-E}{w_p(\tau)}\right)} \geq 0$ for all $w \leq N$.

Indeed, the more the comparison profit is weak, the more the participating firms are keen to avoid the mandatory regulation scheme, even at higher costs. Notice that $w_p(\tau)$

can be explicitly determined in our specified model since (28) becomes a second degree polynomial equation, the unique positive root of which is:

$$w_p(\tau) = \frac{\left(\sqrt{k}N(a-2b\epsilon)\right)}{\sqrt{a^2k+4b\tau-4b^2k\epsilon^2}}. \tag{30}$$

It follows $w_p(\tau) > w_F$ for $\tau > \underline{\tau} = bk\epsilon^2$. Finally, let us stress that w is purely exogenous so that the profitability condition actually relies on the threat.

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

7.2. The symmetric MNE and the optimal CE:

The optimal CE of the game is given by:

$$pr^*(m=N) = \frac{\left(\frac{a^2k}{4b} - bk\epsilon^2 + \tau\right)w - \frac{k}{b}\frac{N^2(a-2b\epsilon)^2}{4w}}{\left(\left(\frac{a^2k}{4b} - bk\epsilon^2 + \tau\right)w - \frac{k}{b}\frac{N^2(a-2b\epsilon)^2}{4w}\right) + \left(\frac{k}{b}\left(\frac{(a-2b\epsilon)}{2}\right)^2\right)},$$

$$pr^*(m=w) = \frac{\frac{k}{b}\left(\frac{(a-2b\epsilon)}{2}\right)^2}{\left(\left(\frac{a^2k}{4b} - bk\epsilon^2 + \tau\right)w - \frac{k}{b}\frac{N^2(a-2b\epsilon)^2}{4w}\right) + \left(\frac{k}{b}\left(\frac{(a-2b\epsilon)}{2}\right)^2\right)},$$

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

$$\begin{aligned} pr^*(s_i=1) &= pr\left(s_i=1 \cap m=N\right) + pr\left(s_i=1 \cap m=w\right) \\ &= pr^*(m=N) + \left(\frac{w}{N}\right) pr^*(m=w), \end{aligned}$$

$$\begin{aligned} pr^*(s_i=0) &= pr\left(s_i=0 \cap m=N\right) + pr\left(s_i=0 \cap m=w\right) \\ &= \left(\frac{N-w}{N}\right) pr^*(m=w), \end{aligned}$$

while the expected social welfare is given by:

$$SW_C^*(\tau) = \frac{kNw(a-2b\epsilon)^2\left(\frac{a^2k}{4b} + \tau - bk\epsilon^2\right)}{\left(1-\frac{N}{w}\right)kN(a-2b\epsilon)^2 + w\left(a^2k + \frac{\tau}{4b} - \frac{k\epsilon^2}{4}\right)}.$$

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

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

We now turn to the numerical simulations, using the specifications that we have introduced.

7.3. The numerical results

Our results are based on an example with $a = 40$ and $b = 2$, which implies $e^{LF} = 10$ and $\pi^{LF} = 200$. We first study the aggregate payoffs generated by p^N and p^* as functions of the tax threat τ, for $k = 0.1$ (i.e. 'weak' sensitivity to emissions, figure 1a) and $k = 2/3$ (i.e. 'strong' sensitivity to emissions, figure 1b). The analysis is done for an ambitious 80%-target ($\epsilon = 2$) so that we can derive an analytical solution of $Q(\tau)$ for $N = 5$ and $N = 10$. We also calculate the corresponding tax interval $[\underline{\tau}, \overline{\tau}]$, the cost of the pigouvian tax, and $-C(N)/N$:

Sensitivity (k)	Tax range ($\underline{\tau}; \overline{\tau}$)	Optimum ($-C(N)/N$)	Tax burden ($\pi^{PIG} - \pi^{LF}$)
0.67	(5.33; 72)	−85.33	128
0.1	(0.8; 180.8)	−12.8	19.2

. Remark that $-C(N)/N$ (black dashed lines in figures 1a and 1b) does not depend on N in this framework since the target is expressed in terms of individual and equally shared burden, ϵ. It is also worth noting that $SW_C^*(\tau)$ depends linearly on the total number of player so that the average payoff per player (blue lines) is the same for all N under p^*. The difference between the red ($SW_N(\tau)/N$ when $N = 5$) and the green ($SW_N(\tau)/N$ when $N = 10$) lines illustrates the 'greed' effect, which is weaker as τ increases and induces greater $NQ(\tau)$. An increase of participation under p^N does not necessarily involve higher expected aggregate payoffs though, since a more stringent τ also increases agents' costs in case of failure, $\sigma(\tau)$:

	Tax min	Average tax	Tax max
N	0.8	90.8	180.8
5	1	0.01	0.003
10	1	0.041	0.014

(a) Weak sensitivity ($k = 0.1$)

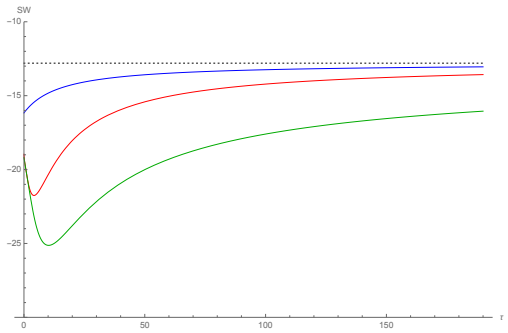
	Tax min	Average tax	Tax max
N	5.33	38.67	72
5	1	0.645	0.351
10	1	0.917	0.669

(b) Strong sensitivity ($k = 0.67$)

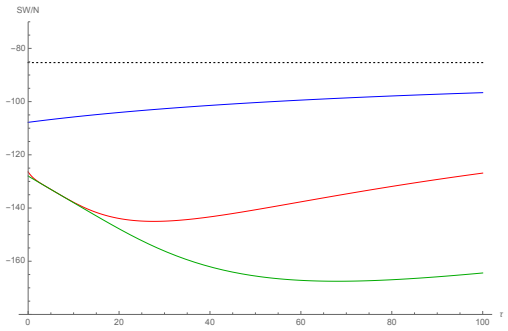
Table 1: Probability of failure of the non-mediated VA under MNE distribution $\sigma(\tau, N)$; with parameter values set at $a = 40$, $b = 2$, and $\epsilon = 2$.

. The decreasing parts of $SW_C^*(\tau)$ in both cases illustrate proposition 5. Specifically, note that in figure 1b, the tax threat undermines the efficiency of p^N as long as $t_0 < 30 + 128$ for $N = 5$ and $t_0 < 70 + 128$ for $N = 10$. Moreover, $t_1 > \overline{\tau} + 128$, which implies that policies $(p^N, \overline{\tau} = 72)$ are Pareto-dominated by $(p^N, 0)$. Implementing a tax threat τ under p^N is therefore not relevant for such parameters values (i.e. high k and low ϵ), especially when the players are numerous. A general normative conclusion may be that when profits are highly sensitive to emissions, using the optimized correlation device p^* remains a significantly more efficient alternative than p^N, even at the maximal tax threat. This result holds for ϵ given, but the next numerical analysis will allow us, inter alia, to

Figure 1: Average expected payoffs per player for $a = 40$, $b = 2$, and $\epsilon = 2$. ${}^{SW^*_C(\tau)}/_N$ is represented by blue lines, and ${}^{SW_N(\tau)}/_N$ by green (resp. red) lines when $N = 10$ (resp. $N = 5$).



(a) $k = {}^1/_{10}$, $\underline{\tau} = 0.8$ and $\overline{\tau} \simeq 180$.



(b) $k = {}^2/_3$, $\underline{\tau} \simeq 5.5$ and $\overline{\tau} = 72$.

state that the adverse effect of a higher τ on $SW_{NE}(\tau)$ weakens with a higher (i.e. less constraining) ϵ.

For a focus on participation rates of distributions $p^*(\tau)$ and $p^N(\tau)$ as well as related sensitivity analysis, see Appendix D. Calculations are done for different target values $\epsilon \in [0, 10]$ and for $N = 10$ and $N = 50$. The case $N = 50 > 30$ is treated as discrete but could be approximated by the normal distribution.

8. Concluding remarks

We have developed a subscription game with a payoff structure representing agents' participation incentives under a preemptive 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 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 mediated VA may succeed. Specifically, we characterize the unique symmetric MNE of the base game and show that it can be implemented by the third party by using the correlation device. Then, we find that the device not only solves the problem raised by multiplicity, but also ensures efficiency gains. Indeed, we characterize the optimal CE of the mediated 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). Lastly, we want to stress that the concept of correlated strategies applied to such a public good provision game may turn out to be of particular interest when considering license plate-based driving restrictions.

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

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

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

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

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

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

$$x_1 = pr(m = w-1)\left[\left(\frac{(N-w)C(N)}{C(w)-tw-C(N)}\left(\frac{C(w+1)}{w+1}\right)\right) + (N-(w-1))\left(\frac{C(w)-tw}{w}\right)\right]$$
$$+ \sum_{k=w+1}^{N-1} pr(m = k)\left(\frac{(N-k)(C(k+1))}{k+1} + \left(\frac{(C(N)-C(k))}{C(w)-tw-C(N)}\left(\frac{(N-w)C(w+1)}{w+1}\right)\right)\right)$$
$$-\frac{(N-w)C(N)}{C(w)-tw-C(N)}\left(\frac{C(w+1)}{w+1}\right) - \frac{x_2(N-w)}{C(w)-tw-C(N)}\left(\frac{C(w+1)}{w+1}\right)$$
$$+ \sum_{k=0}^{w-2} pr(m = k)\left(\frac{(N-w)C(N)}{C(w)-tw-C(N)}\left(\frac{C(w+1)}{w+1}\right)\right)$$

$$pr(m = w) = \sum_{k=0}^{w-1} pr(m = k)\left(\frac{C(N)}{C(w)-tw-C(N)}\right) - \frac{C(N)}{C(w)-tw-C(N)}$$
$$+ \sum_{k=w+1}^{N-1} pr(m = k)\left(\frac{C(N)-C(k)}{C(w)-tw-C(N)}\right) - \frac{x_2}{C(w)-tw-C(N)} \tag{B.1}$$

$$pr\left(m=N\right)=1-\sum_{k=0}^{w-1}pr\left(m=k\right)\left(\frac{C(w)-tw}{C(w)-tw-C\left(N\right)}\right)+\frac{C\left(N\right)}{C(w)-tw-C\left(N\right)}$$
$$-\sum_{k=w+1}^{N-1}pr\left(m=k\right)\left(\frac{C(w)-tw-C\left(k\right)}{C(w)-tw-C\left(N\right)}\right)+\frac{x_2}{C(w)-tw-C\left(N\right)} \tag{B.2}$$

We substitute (B.1) and (B.2) into the objective (17). It follows the objective depends only on the N non-basic variables :

$$\left[\sum_{k=0}^{w-1}pr\left(m=k\right)\left(\left(\frac{C\left(N\right)\left(C(w)-tw\right)}{C(w)-tw-C\left(N\right)}\right)-\left(\frac{C\left(w\right)C\left(N\right)}{C(w)-tw-C\left(N\right)}\right)-Nt\right)\right]$$
$$+\left[\sum_{k=w+1}^{N-1}pr\left(m=k\right)\left(C\left(N\right)\left(\frac{C(w)-tw-C\left(k\right)}{C(w)-tw-C\left(N\right)}\right)-C\left(w\right)\left(\frac{C\left(N\right)-C\left(k\right)}{C(w)-tw-C\left(N\right)}\right)-C\left(k\right)\right)\right]$$
$$+C\left(N\right)\left(\frac{tw}{C(w)-tw-C\left(N\right)}\right)+\left(\frac{C\left(w\right)-C\left(N\right)}{C(w)-tw-C\left(N\right)}\right)x_2 \tag{B.3}$$

Rearranging (B.3) we obtain :

$$\sum_{k=0}^{w-1}pr\left(m=k\right)\left(\frac{-C\left(N\right)tw}{C(w)-tw-C\left(N\right)}-Nt\right)+C\left(N\right)\left(\frac{tw}{C(w)-tw-C\left(N\right)}\right) \tag{B.4}$$
$$+\sum_{k=w+1}^{N-1}pr\left(m=k\right)\left(\frac{tw\left(C\left(k\right)-C\left(N\right)\right)}{C(w)-tw-C\left(N\right)}\right)+\left(\frac{C\left(w\right)-C\left(N\right)}{C(w)-tw-C\left(N\right)}\right)x_2 \tag{B.5}$$

Then, let us remark that:

$$\underbrace{\left(\frac{C\left(N\right)}{C\left(N\right)+tw-C(w)}\right)}_{<1}w<N. \tag{B.6}$$

Coefficients (dual variables) associated to $\{pr(m=k)\}_0^{w-1}$ are therefore negative. Assumptions on cost and profitability imply the remaining components of (B.5) are 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:

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

$$\Rightarrow\begin{cases}x_1^* & =\frac{(N-w)C(N)}{(tw-C(w))+C(N)}\frac{C(w+1)}{w+1}\\ pr^*(m=w) & =\frac{C(N)}{(tw-C(w))+C(N)}\\ pr^*(m=N) & =\frac{(tw-C(w))}{(tw-C(w))+C(N)}.\end{cases} \tag{B.7}$$

Appendix C. Proof of Proposition 3

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

$$pr\left(m=k\mid s_i=1\right)=\frac{pr\left(m=k\right)pr\left(s_i=1\mid m=k\right)}{pr\left(s_i=1\right)}=\left(\left(\begin{array}{c}N\\k\end{array}\right)q^k(1-q)^{N-k}\right)\frac{k}{qN}$$

$$pr\left(m=k \mid s_i=0\right)=\frac{pr\left(m=k\right)pr\left(s_i=0 \mid m=k\right)}{pr\left(s_i=0\right)}=\left(\left(\begin{array}{c} N \\ k \end{array}\right)q^k(1-q)^{N-k}\right)\frac{N-k}{N}\frac{1}{1-q},$$

Now, using the two previous equalities, we can rewrite condition (14) of CE as follows:

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

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

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

Likewise, condition (15) of CE rewrites:

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

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

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

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

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

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

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

which is the symmetrical MNE condition when m=j=0. It follows the unique MNE is a symmetric CE.

Appendix D. Proof of Proposition 5

First, remark that the probability constraint involves the derivative of $SW_{NE}(t)$ with respect to t can be rearranged as follows:

$$-\frac{d}{dt}\sum_{w}^{\lceil NQ(t)\rceil-1}pr(m=k)\left(\frac{C\left(k\right)}{N}-t\right)-\frac{d}{dt}\sum_{\lceil NQ(t)\rceil}^{N}pr(m=k)\left(\frac{C\left(k\right)}{N}+t\right)-\sum_{0}^{w}pr(m=k), \qquad \text{(D.1)}$$

since $\frac{d}{dt}\sum_{0}^{w-1}pr(m=k)+\frac{d}{dt}\sum_{w}^{\lceil NQ(t)\rceil-1}pr(m=k)=\frac{d}{dt}\sum_{\lceil NQ(t)\rceil}^{N}pr(m=k)$. As m obviously follows the binomial $\mathcal{B}(N,Q(t))$, it can be approximated by the normal distribution $\mathcal{N}(0,1)$, with change of

variable: $M = \frac{m-NQ(t)}{\sqrt{NQ(t)(1-Q(t))}}$. The derivative (D.1) therefore rewrites :

$$\int_{\frac{w-NQ(t)}{\sqrt{NQ(t)(1-Q(t))}}}^{0} \frac{1}{2\pi} e^{-\frac{x^2}{2}} dx \underbrace{- \int_{0}^{\frac{N-NQ(t)}{\sqrt{NQ(t)(1-Q(t))}}} \frac{1}{2\pi} e^{-\frac{x^2}{2}} dx - \int_{\frac{-NQ(t)}{\sqrt{NQ(t)(1-Q(t))}}}^{\frac{w-NQ(t)}{\sqrt{NQ(t)(1-Q(t))}}} \frac{1}{2\pi} e^{-\frac{x^2}{2}} dx}_{:=A} \tag{D.2a}$$

$$+ \frac{1}{2\pi} e^{-\frac{\left(\frac{w-NQ(t)}{\sqrt{NQ(t)(1-Q(t))}}\right)^2}{2}} \left(t - \frac{C(w)}{N}\right) \left(\frac{N\left(w + Q(t)\left(N - 2w\right)\right)}{2\left(NQ(t)\left(1-Q(t)\right)^{3/2}\right)}\right) \tag{D.2b}$$

$$+ \frac{1}{2\pi} e^{-\frac{\left(\frac{N-NQ(t)}{\sqrt{NQ(t)(1-Q(t))}}\right)^2}{2}} \left(\frac{C(N)}{N} + t\right) \left(\frac{N}{2Q(t)\sqrt{NQ(t)\left(1-Q(t)\right)}}\right) \tag{D.2c}$$

Assuming $t > t_w$ and $C'(m) < 0$, we know the terms (D.2b) and (D.2c) are positive. Finally, as $Q(t)$ increases in t, and $w - NQ(t) < 0 < N - NQ(t)$ for all $Q(t) > 0$, we know the first term (respectively, the second term) of (D.2a) increases (respectively, decreases) in t. It follows there exists t_0 such that $\int_{\frac{w-NQ(t_0)}{\sqrt{NQ(t_0)(1-Q(t_0))}}}^{0} \frac{1}{2\pi} e^{-\frac{x^2}{2}} dx = -A$, as well as $t_1 < t_0$ from which the whole expression becomes positive.

Appendix E. 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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 E.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 (E.8b), (E.8d) and (E.7), we get the set of candidates $\left(p_{11}\ p_{10}\ p_{01}\ 0\right)^T \in \mathbb{R}^4$ such that

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

Geometricaly, it is the area bounded by the inequalities

$$p_{10} \geq \frac{C(2)}{2\left(t-c\right)+C(2)} - \left(\frac{C(2)}{2\left(t-c\right)+C(2)}\right)p_{01} \tag{E.11}$$
$$p_{10} \geq 1-\left(1+\frac{2\left(t-c\right)}{C(2)}\right)p_{01} \tag{E.12}$$
$$p_{10} \leq 1-p_{01} \tag{E.13}$$

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

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

as illustrated in figure E.2. 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

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

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

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

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

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

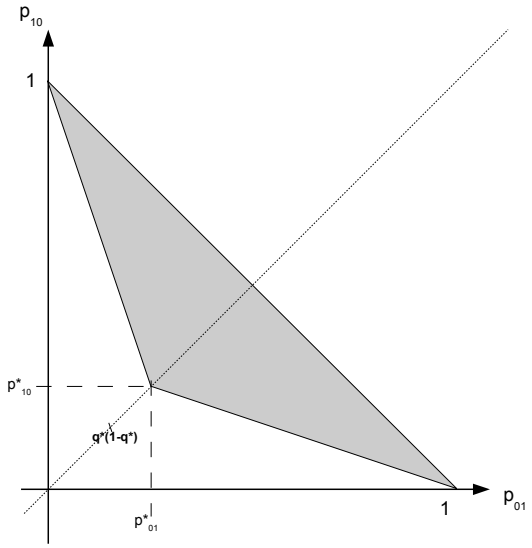


Figure E.2: The set of correlated equilibria in the two-player example.

Or, geometrically, the grey area extends along the bisectrice on figure E.2. The aggregate payoff tends then to the payoff corresponding to the socially optimal allocation

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

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 (E.8a)-(E.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

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

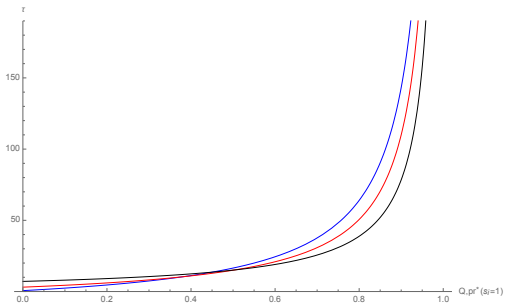
implying that the coordination device not only solve the problem raised by multiplicity, but also ensures that a higher expected aggregate payoff is reached for a given credible level of threat.

Appendix F. Numerical simulations

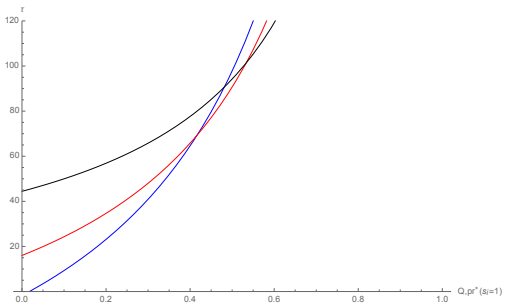
. Our results are summarized in figures F.3a to F.4d. Calculations are done for $N = 10$ and $N = 50$ The case $N = 50 > 30$ is treated as discrete but could be approximated by the normal distribution.

. A first (trivial) comment regards the impact of τ on participation decisions. Figures F.3a to F.4d show that both $\tau(Q)$ and $\tau(pr^*(s_i = 1))$ increase in participation rates, i.e. in terms of reciprocal: the stronger the threat is, the more the players are prone to participate at the equilibrium. Numerical results are consistent with general eq. (11), and (19), the examination of which predicts the effect of a tax threat variation is direct and unambiguous since only payoffs in case of failure are affected. Comparing F.3a with F.3b, and F.3c with F.3d, as well as F.4a with F.4b, and F.4c with F.4d, we see this effect decreases in k for both distributions. Then, focus on eq. (19), and remark that the impact of an increase of the

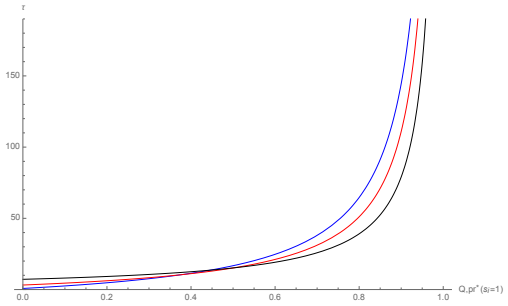
Figure F.3: Participation rates under p^* for the target values $\epsilon = 2$ (blue), $\epsilon = 4$ (red) and $\epsilon = 6$ (black).



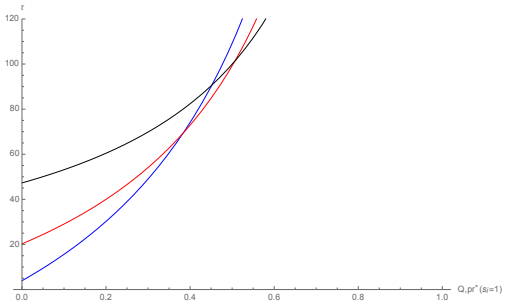
(a) $k = 1/10$, $N = 10$



(b) $k = 2/3$, $N = 10$



(c) $k = 1/10$, $N = 50$



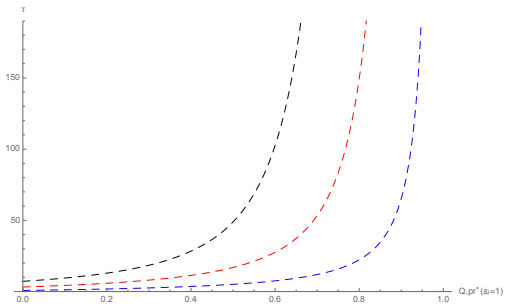
(d) $k = 2/3$, $N = 50$

target on $pr^*(s_i = 1)$ is twofold : on one hand, it cuts the cost to participate for all m (direct effect on $C(m)/m = (N(\epsilon-10)/5m)^2$), on the other hand it makes the level of feasible w smaller with factor t (indirect effect via $w_F = N(1-\epsilon/10)$), which reduces, in turn, $pr^*(s_i = 1)$. The indirect effect dominates the direct effect when t becomes high enough. This occurs, graphically, when plain lines intersect in figures F.3a to F.3d. Finally, dashed lines show that a less ambitious target unambiguously increases $\tau(Q)$. Indeed, the pivotality rate under p^N, which composes $\tau(Q)$ as defined by proposition 1, rewrites:

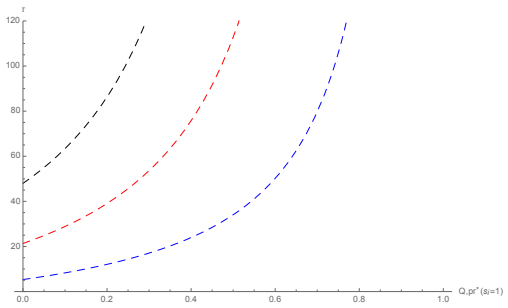
$$\sum_{k>w}^{N} \frac{pr(m = k \mid s_i = 1)}{pr(m = w \mid s_i = 1)} = \sum_{k>\lceil N(1-\epsilon/10)\rceil}^{N} \frac{k}{\lceil N(1-\epsilon/10)\rceil} \frac{\binom{N}{k}}{\binom{N}{\lceil N(1-\epsilon/10)\rceil}} \left(\frac{q}{1-q}\right)^{k-\lceil N(1-\epsilon/10)\rceil}$$

when $w = \lceil N(1-\epsilon/10)\rceil$. It therefore increases in ϵ via w (indirect effect), while participation costs for any given number of participants decrease. Finally, we point out that when $k = 1/10$ and the target is set equal to 2, a tax threat $\tau > 68$ turns out to yield higher participation rates under p^N (figure F.4c) than under p^* (figure F.3c). However, figure 1a accordingly with previous section's theoretical results, show that p^N always generates smaller aggregate payoffs, even when exhibiting higher participation rates. Let us conclude this numerical analysis with a figure which precisely represents the difference in participation rates under p^* and p^N for $N = 10$ and $N = 50$. We see the difference increases in N when $\tau(pr^*(s_i = 1)) - \tau(Q) < 0$. Indeed, $\tau(Q)$ relies on a pivotality rate that increases (directly and indirectly via w) in N, which means more free-riding and therefore higher τ for any given Q, while $pr^*(m = w)$ and $pr^*(m = N)$ which define the specified optimal CE do not depend on the total number of player (and, as a result, $SW^*_C(\tau)$ does not either, as illustrated in figure 1a and 1b).

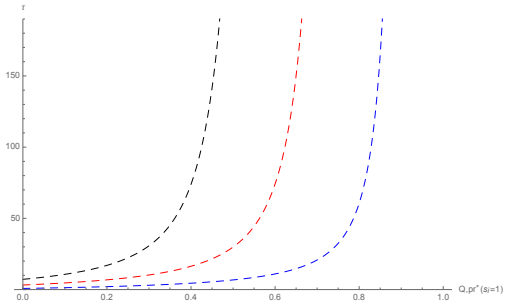
Figure F.4: Participation rates under p^N for the target values $\epsilon = 2$ (blue), $\epsilon = 4$ (red) and $\epsilon = 6$ (black).



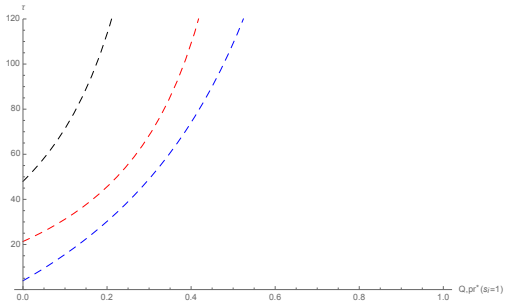
(a) $k = 0.1$, $N = 10$



(b) $k = 0.67$, $N = 10$

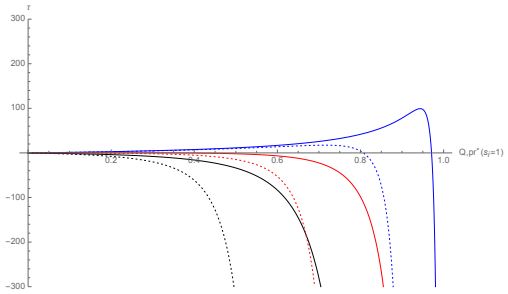


(c) $k = 0.1$, $N = 50$

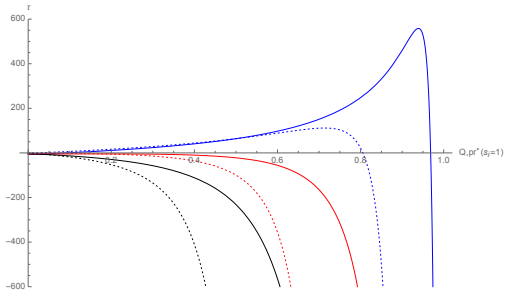


(d) $k = 0.67$, $N = 50$

Figure F.5: The gap between participation rates := $\tau(pr^*(s_i = 1)) - \tau(Q)$, when the total number of player is $N = 10$ (plain lines) and N=50 (dotted lines), for $\epsilon = 2$ (blue), $\epsilon = 4$ (red) and $\epsilon = 6$ (black).



(a) $k = 1/10$



(b) $k = 2/3$

References

[1] Arce, D. G. and Sandler, T. (2001). Transnational public goods: strategies and institutions. *European Journal of Political Economy*, 17:493–516. 3

[2] Aumann, R. J. (1974). Subjectivity and correlation in randomized strategies. *Journal of Mathematical Economics*, 1(1):67–96. 2, 5, 13

[3] Borkey, P., Glachant, M., and Leveque, F. (1998). Voluntary approaches for environmental policy in oedc countries: An assessment. Technical report, Paris: CERNA, Centre d'économie Industrielle. 2

[4] Brau, R. and Carraro, C. (2011). The design of voluntary agreements in oligopolistic markets. *Journal of Regulatory Economics*, 39(2):111 – 142. 2, 20

[5] Cavaliere, A. (2001). Coordination and the provision of discrete public goods by correlated equilibria. *Journal of Public Economic Theory*. 3

[6] 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. 15

[7] d'Aspremont, C., Jacquemin, A., Gabszewicz, J. J., and Weymark, J. A. (1983). On the stability of collusive price leadership. *Canadian Journal of Economics*, 16(1):17–25. 4

[8] Dawson, N. L. and Segerson, K. (2008). Voluntary agreements with industries: Participation incentives with industry-wide targets. *Land Economics*, 84(1):97–114. 2, 4

[9] Glachant, M. (2007). Non-binding voluntary agreements. *Journal of Environmental Economics and Management*, 54(1):32–48. 2

[10] Myerson, R. B. (1997). *Game Theory Analysis of Conflict*, chapter 6. Harvard University Press. 10, 27

[11] Nisan, N., Roughgarden, T., Tardos, E., and Vijay, V. (2007). *Algorithmic Game Theory*. Cambridge University Press. 13

[12] OECD (1999). Voluntary approaches for environmental policy : An assessment. 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, 9, 22, 23

[14] Segerson, K. and Wu, J. (2006). Nonpoint pollution control: Inducing first-best outcomes through the use of threats. *Journal of Environmental Economics and Management*, 51(2):165–184. 2

[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. 15