UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Scienze Economiche “Marco Fanno”

ON THE ECONOMIC VALUE OF REPEATED INTERACTIONS UNDER ADVERSE SELECTION

OTTORINO CHILLEMI
Università di Padova

BENEDETTO GUI
Università di Padova

LORENZO ROCCO
Università di Padova

October 2005

“MARCO FANNO” WORKING PAPER N.7

On the economic value of repeated interactions under adverse selection

Ottorino Chillemi*, Benedetto Gui† and Lorenzo Rocco‡
University of Padua

October, 2005

Abstract

The paper studies, in a repeated interaction setting, how the presence of cooperative agents in a heterogeneous community organized in groups, affects group efficiency and stability. The paper extends the literature by assuming that each type can profitably mimic other types. It is shown that such enlargement of profitable options prevents group stabilization in the single group case. Stabilization can be obtained with many groups, but its driver is not the efficiency gain due to the presence of cooperative individuals. Instead stabilization is the result of free riding opportunities.

1 Introduction

This paper studies how the degree of cooperativeness affects efficiency in a community organized in groups, when individual type is private information and group stability is endogenous. In this regards our interest is akin to Ghosh and Ray's (1996). They consider a population composed by patient and impatient individuals. Initially, individuals are randomly matched in pairs to play a prisoner dilemma in continuous strategies. At the end of each period partners may opt between continuing their current relationship or separating and trying another random matching. In case of break off, information about partners' type is not disclosed to the rest of the population. It is shown that the presence of the impatient type gives a value to the fact of being matched with the good type, which can support cooperation among patient players. Equilibrium strategies contemplate an initial round in which players test their fellows by setting a moderate level of cooperation. If the partner is patient, he does the same and an ever-lasting relationship starts, where cooperation is at its highest possible level. If the partner is impatient, he does

*Dipartimento di Scienze Economiche,Università di Padova, Via del Santo 33 . tel: +39 49 8274224; e-mail: ottorino.chillemi@unipd.it
†Dipartimento di Scienze Economiche,Università di Padova, Via del Santo 33 . tel: +39 49 8274225; e-mail: benedetto.gui@unipd.it
‡Dipartimento di Scienze Economiche,Università di Padova, Via del Santo 33 . tel: +39 49 8274260; e-mail: lorenzo.rocco@unipd.it

not cooperate and the relationship breaks down. Therefore, equilibrium permanent relationships emerge with a positive effect on social surplus. Similar result can be found in Kranton (1996) – for later contributions see Rauch and Watson (2002) and Watson (2002) – and in the reputation building literature following the seminal paper of Milgrom and Roberts (1982).

At first sight, Ghosh and Ray's results are somewhat surprising, since one might expect that the incentive costs of exploiting information, elicited in the early phase of a relationship, exceed the benefit from exploitation itself, according to the logic of the "ratchet" effect. However, a careful examination shows that such exploitation costs never arise in their framework. On the one hand, "bad" (impatient) agents do not face any temptation to mimic patient players - they always find it advantageous to defect at the first occasion. On the other hand, "good" agents can only accede to the benefits of a stable relationship by being considered patient. So, absent any incentive to dissimulate one's true type, the ratchet effect is excluded by definition.

Instead, in the model we present below the bad type may mimic the good type for the relationship to be stabilized and then, when stabilization will finally obtain, exploit their partners - we keep the assumption that information disclosure is local.

Let us briefly describe our setting. We consider a two-period model in which members of a large population of risk-neutral individuals interact in groups. Initially, groups are formed randomly. After completing the first round of interaction, each group can either continue interaction or dissolve. In case of dissolution, former group members are randomly assigned to newly formed "fresh" groups. In both periods group members bargain over sharing the cost of provision of an indivisible local public good. Bargaining is modelled as a direct revelation mechanism in which a benevolent principal maximizes expected group surplus over the rest of the game[1]. Individual type is bi-dimensional. The first dimension is the degree of alignment of the objective function of the agent with that of the principal (we refer to this feature as altruism towards fellow group members). This is a time-invariant characteristic, which is privately learned at the beginning of the whole game. The second dimension, which is privately learned at the beginning of each period, is agent's own "material" benefit deriving from the consumption of the public good. The probability distribution of types is the same across individuals and also over time, and is common knowledge. Moreover, the two components of an individual's type are stochastically independent. We also assume that no intertemporal and intergroup transfers are possible and that information revealed within a group does not leak outside.

[1] Group payoff thus obtained is to be viewed as an upper bound to the expected group surplus that can be obtained through any bargaining procedure.

Decisions concerning continuation/stabilization of a group are taken by the first-period principal. In case of continuation he transmits information about group composition to group members and to his second-stage successor, whose task reduces to operating a mechanism for the revelation of agents' second-period levels of material benefit from the public good. We analyze the adverse incentive effects that prevent group members from taking full advantage from the information about group composition that emerges in the first stage of the game. We show that stable relationships can emerge in equilibrium; this has positive value for group members, but hurts society, in the sense that social surplus is lower than in a situation in which stable relationships are forbidden.

The rest of the paper is organized as follows. Section 2 presents the model, Section 3 contains the main results, Section 4 discusses the robustness of our findings and suggests possible extensions, Sections 5 concludes. Minor technical proofs are collected in two appendices.

2 The model

At date $t = 1$, nature randomly forms N groups, each one composed of n members. Next in each group a benevolent surplus maximizer principal announces the probabilities of provision of the public good and the payments due by each individual. Both probabilities and payments are functions of all members' reported types. The principal also announces the probability that the group is stabilized, that is, all members remain together, in which case he will inform the second-period principal and all members about individuals' permanent types, as a function of reported types. Observing the menu of proposed allocations and probabilities of group stabilization, each group member confidentially reports his type to his principal. At date $t = 2$, non stabilized groups dissolve and their members are randomly matched into fresh groups, where they play again the direct revelation mechanism faced at date 1, under the rule of a new principal – of course, the decision about group's stabilization is inessential in the second period since as it ends the game stops. Instead, the members of stabilized groups play a direct mechanism for the revelation of their material benefit under the supervision of a new principal and common knowledge of members' permanent types. For the sake of simplicity, in our analysis agents learn about their mates' types by observing realized payments. To this end, in the model goods are produced and consumed, and transfers are paid only and the end of the second period.

The sake of simplicity also motivates the assumption that the first-period and the second-period principals do not coincide. This allows us to avoid consideration of intertemporal strategies and also of the tricky issue of what information the principal should communicate to the agents[2].

Agents' preferences are modelled as follows. Type-e agents maximize their own present and future material benefits; type-a agents are more aligned with the principal: they assign a positive weight also to the material benefit (present and future) of the current fellows, provided that their own material benefit is positive – an agent's utility being infinitely negative otherwise.

Therefore type-a agents do not take into account the benefit that will accrue to the new fellows they will interact with in case the group is dissolved [3]. However, if dissolution occurs, when taking their second period decisions, they will be concerned with the welfare of their fellows at that time. Furthermore, from the outset, they rationally anticipate the consequences of such concern on their own future material benefits. This kind of concern characterizes the attitude of an agent who, when interacting with a group of people, takes his fellows' welfare into account. However, their welfare is not for him a source of utility to be included in his own intertemporal maximand (on this see Sen 1985).

In order to set a reasonable limit to the possibility of redistribution, we posit that individual payments cannot be negative. Furthermore we posit that the budget must balance within each group in each period.

To further simplify, we will only consider the cases in which members of a first-period group meet again in fresh groups with probability one and with probability zero, respectively – so we disregard the case in which this probability is fractional. This corresponds to $N = 1$ (one group only), and to $N = \infty$ (a very large number of groups), respectively.

Given that individuals are ex-ante identical, we can limit our attention to anonymous mechanisms; moreover, we focus on truth-telling, symmetric, perfect Bayes-Nash equilibria.

Now, we can state a list of objects.

The individual type in period i is $\mathbf{t}^i = (t_1, t_2^i) \in T = \{a, e\} \times \{h, l\}$ where t_1 is the alignment parameter, assumed to be time invariant, and t_2^i is the private benefit from the public good, a time-varying parameter.

The cost of the public good is $c > 0$.

At date 1, it is common knowledge that $\Pr(t_1 = a) = \theta \in (0, 1)$, $\Pr(t_1 = e) = 1 - \theta$, $\Pr(t_2^i = h) = p \in (0, 1)$, $\Pr(t_2^i = l) = 1 - p$, $\Pr(\mathbf{t}^i) = \Pr(t_1) \Pr(t_2^i)$, all i.

A realization of agents' type profile, i.e. a state, at date i is denoted $\mathbf{v}^i$. A state is

[2]The timing of production and payment can be the usual one – that is, public goods are produced and transfers are paid at the end of each period – when the number of active groups is very large, since in such a case the realized state in a first period group does not convey any additional information on the population distribution.

[3]Instead, in his first-period decisions a type-a agent assigns a positive weight to the future welfare of his first-period fellows even if in the second period they will be dispersed.

completely described by the number of agents of each type, $v^i_{\mathbf{t}^i}$, all $\mathbf{t}^i$, so $\mathbf{v}^i$ is defined as $\mathbf{v}^i = \left(v^i_{ah}, v^i_{al}, v^i_{eh}, v^i_{el}\right)$. Heretofore we will omit the zeros and - for the reader's comfort – we write, for example, $\mathbf{v}^i = (nah)$ in place of $\mathbf{v}^i = (n, 0, 0, 0)$, $\mathbf{v}^i = (kah.(n-k)eh)$ in place of $\mathbf{v}^i = (k, 0, n-k, 0)$, and so on.

When considering the possibility of agents deviating from truth-telling, we need to consider only one deviator at a time. So we express the reported state as a function of the true state and the deviation in the following way.

Let $j \in J = \{1, 2, s_k,\ k = 0, 1, .., n\}$ denote, respectively, a period-one group, a period-two fresh group, a period-two stabilized group, with k type-a members. Let $\mathbf{d}^j \in T$ denote the type reported in group j by the agent under consideration; then $\widehat{\mathbf{v}}(\mathbf{v}^i, \mathbf{t}^i, \mathbf{d}^j)$ is the reported state in group j when the true state is $\mathbf{v}^i$ (recall that states and types are time-depend), the reporter is of type $\mathbf{t}^i$ and declares $\mathbf{d}^j$ while all the other individuals truthfully report their types. Hereafter, for simplicity we omit all the arguments of $\widehat{\mathbf{v}}$ and refer to the kind of group under examination by adding the apex j: i.e. $\widehat{\mathbf{v}}^j$ denotes the reported state in group j. If $\mathbf{t}^i = \mathbf{d}^j$, obviously $\widehat{\mathbf{v}}^j = \mathbf{v}^i$; if instead $\mathbf{d}^j \neq \mathbf{t}^i$, it is $\widehat{\mathbf{v}}^j = \{\widehat{v}^j_{\boldsymbol{\tau}} : \widehat{v}^j_{\boldsymbol{\tau}} = v^i_{\boldsymbol{\tau}}$ for $\boldsymbol{\tau} \notin \{\mathbf{t}^i, \mathbf{d}^j\}; \widehat{v}^j_{\boldsymbol{\tau}} = v^i_{\boldsymbol{\tau}} - 1$ for $\boldsymbol{\tau} = \mathbf{t}^i$, $\widehat{v}^j_{\boldsymbol{\tau}} = v^i_{\boldsymbol{\tau}} + 1$ for $\boldsymbol{\tau} = \mathbf{d}^j\}$. To emphasize that in second period stabilized groups the permanent type has already been revealed at time 1, we write $\mathbf{d}^j = \left(d^1_1, d^j_2\right)$.

Now, we start to define a second list of more complex objects.

The payment, if any, due by an individual who reports $\mathbf{d}^j$ in a group j when the reported state is $\widehat{\mathbf{v}}^j$ is denoted $g_j(\widehat{\mathbf{v}}^j, \mathbf{d}^j)$. The probability of provision of the public good when the reported state is $\widehat{\mathbf{v}}^j$ is denoted $r_j(\widehat{\mathbf{v}}^j) \in [0, 1]$. Moreover, $\sigma\left(\widehat{v}^1_{ah} + \widehat{v}^1_{al}\right)$ is the probability that a first-period principal opts for stabilization and informs the second-period principal about individual types: it is a function of the reported number of type-a members. We denote $\Psi_j = \left\{g_j(\widehat{\mathbf{v}}^j), r_j(\widehat{\mathbf{v}}^j), \boldsymbol{\sigma}\left(\widehat{v}^j_{ah} + \widehat{v}^j_{al}\right),\ \text{all } \widehat{\mathbf{v}}^j,\ \text{all } \widehat{v}^j_{ah} + \widehat{v}^j_{al}\right\}$ the strategy of a j-group principal, all j, where $\boldsymbol{\sigma}\left(\widehat{v}^j_{ah} + \widehat{v}^j_{al}\right)$ is ineffective for $j \neq 1$.

The material surplus, if any, accruing in group j to an individual of type $\mathbf{t}^i$ who reports $\mathbf{d}^j$ where $\widehat{\mathbf{v}}^j$ is the reported state is $s_j(\mathbf{t}^i, \widehat{\mathbf{v}}^j, \mathbf{d}^j)$. We define

$$s_j(\mathbf{t}^i, \widehat{\mathbf{v}}^j, \mathbf{d}^j) = d^j_2 - g_j(\widehat{\mathbf{v}}^j, \mathbf{d}^j) \tag{1}$$

The unconditional probability of state $\mathbf{v}^i$ is $\Pr\left(\mathbf{v}^i\right)$ while $\Pr_j\left(\mathbf{v}^i\right)$ denotes the probability of state $\mathbf{v}^i$ as seen by a j-group principal. To illustrate the latter, let $n = 2$. Then, for instance, $\Pr_{s1}(ah.eh) = p^2$; $\Pr_1(ah.eh) = 2\theta p(1-\theta)p$; $\Pr_2(ah.eh) = 2\Theta p(1-\Theta)p$, where Θ is obtained by applying Bayes' rule using the strategies of first-period principals. Moreover, $\Pr(k)$ denotes the probability a first-period principal assigns to the event that the number of type-a members in his group is k.

Under truth-telling, the principal of a j-group computes current expected surplus as

$$\Phi_j = \sum_{\mathbf{v}^i \in V} \Pr_j(\mathbf{v}^i) r_j(\mathbf{v}^i) \sum_{\boldsymbol{\tau} \in T: v_{\boldsymbol{\tau}} > 0} v^i_{\boldsymbol{\tau}} s_j(\boldsymbol{\tau}, \mathbf{v}^i, \boldsymbol{\tau}) \tag{2}$$

(Notice that $i = 1$ when $j = 1$ and $i = 2$ otherwise.)

>From the perspective of a member of type $\mathbf{t}^1$ who reports $\mathbf{d}^1$, his first-period expected material surplus is

$$z_1\left(\mathbf{t}^1, \mathbf{d}^1\right) = \sum_{\mathbf{v}^1 \in V} \Pr_1(\mathbf{v}^1 |\ \mathbf{t}^1) r_1(\widehat{\mathbf{v}}^1) s_1(\mathbf{t}^1, \widehat{\mathbf{v}}^1, \mathbf{d}^1) \tag{3}$$

Similarly, the expected surplus in a j-group, $j \neq 1$, of an agent of type $\mathbf{t}^2$ who reported d^1_1 in the first period and reports $\mathbf{d}^j$ in the second period is

$$z_j\left(\mathbf{t}^2, d^1_1, \mathbf{d}^j\right) = \sum_{\mathbf{v}^2 \in V} \Pr_j(\mathbf{v}^2 |\ \mathbf{t}^2, d^1_1) r_j(\widehat{\mathbf{v}}^j) s_j(\mathbf{t}^2, \widehat{\mathbf{v}}^j, \mathbf{d}^j) \tag{4}$$

Moreover, an individual who is of type $\boldsymbol{\tau}^2$ and reports $\mathbf{d}^2$ in the second period after having reported d^1_1 in the first period, will compute the surplus in a fresh group to a type $\mathbf{t}^2$ individual as follows

$$z^{\mathbf{t}^2}_2\left(\boldsymbol{\tau}^2, d^1_1, \mathbf{d}^2\right) = \sum_{\mathbf{v}^2 \in V} \Pr_2(\mathbf{v}^2 |\ \boldsymbol{\tau}^2, d^1_1) r_2(\widehat{\mathbf{v}}^2) s_2(\mathbf{t}^2, \widehat{\mathbf{v}}^2, \mathbf{t}^2) \tag{5}$$

In the first period, a type $\mathbf{t}^1$ individual who reports $\mathbf{d}^1$ expects that the total first period surplus to his fellows is

$$\phi_1\left(\mathbf{t}^1, \mathbf{d}^1\right) = \sum_{\mathbf{v}^1 \in V} \Pr_1(\mathbf{v}^1 |\ \mathbf{t}^1) r_1(\widehat{\mathbf{v}}^1) \left[(\widehat{v}^1_{\mathbf{d}^1} - 1) s_1(\mathbf{d}^1, \widehat{\mathbf{v}}^1, \mathbf{d}^1) + \sum_{\boldsymbol{\tau}^1 \in T, \boldsymbol{\tau}^1 \neq \mathbf{d}^1} \widehat{v}^1_{\boldsymbol{\tau}^1} s(\boldsymbol{\tau}^1, \widehat{\mathbf{v}}^1, \boldsymbol{\tau}^1) \right] \tag{6}$$

Similarly, for $j \neq 1$, in a j-group, a type $\mathbf{t}^2$ individual who reports $\mathbf{d}^j$, after having reported d^1_1 in the first period expects that the total surplus accruing to his current fellows is

$$\phi_j\left(\mathbf{t}^2, d^1_1, \mathbf{d}^j\right) = \sum_{\mathbf{v}^2 \in V} \Pr_j(\mathbf{v}^2 |\ \mathbf{t}^2, d^1_1) r_j(\widehat{\mathbf{v}}^j) \left[(\widehat{v}^j_{\mathbf{d}^j} - 1) s_j(\mathbf{d}^j, \widehat{\mathbf{v}}^j, \mathbf{d}^j) + \sum_{\boldsymbol{\tau}^2 \in T, \boldsymbol{\tau}^2 \neq \mathbf{d}^j} \widehat{v}^j_{\boldsymbol{\tau}^2} s(\boldsymbol{\tau}^2, \widehat{\mathbf{v}}^j, \boldsymbol{\tau}^2) \right] \tag{7}$$

Let $\widehat{k}\left(k, t_1, d^1_1\right)$ denote the number of agents who report to be of type-a in a first-period group, where the actual number of altruists is k and a member of type t_1 reports d^1_1 while his fellows tell the truth.

A type $\boldsymbol{\tau}^2$ agent who reports $\mathbf{d}^2$ in the second period after having reported d_1^1 in the first period expects, conditionally on the number of reported altruists $\widehat{k}$, that the total surplus accruing in fresh groups to his former fellows is

$$\Omega_2(\boldsymbol{\tau}^2, d_1^1, \mathbf{d}^2, \widehat{k}) = \left(\widehat{k} - 1_{[d_1^1=a]}\right) E_{t_2^2} z_2^{\left(a,t_2^2\right)} \left(\boldsymbol{\tau}^2, d_1^1, \mathbf{d}^2\right) + \\ + (n - \widehat{k} - 1 + 1_{[d_1^1=a]}) E_{t_2^2} z_2^{\left(e,t_2^2\right)} (\boldsymbol{\tau}^2, d_1^1, \mathbf{d}^2) \quad (8)$$

where $1_{[d_1^1=a]}$ is a dummy which takes 1 if $d_1^1 = a$ and zero otherwise. Notice that this definition covers the cases $N = 1$ and $N = \infty$ only.

Let $z_j^{\bigstar}(\cdot) = z_j(\cdot)$ if $z_j(\cdot) \geq 0$, and $z_j^{\bigstar}(\cdot) = -\infty$ otherwise, all j. Then, the objective function of an agent in group $j \neq 1$ is

$$U_j\left(\mathbf{t}^2, d_1^1, \mathbf{d}^j\right) = z_j^{\bigstar}\left(\mathbf{t}^2, d_1^1, \mathbf{d}^j\right) + t_1 \phi_j\left(\mathbf{t}^2, d_1^1, \mathbf{d}^j\right) \quad (9)$$

i.e. the sum of individual material surplus and of his fellows' surplus weighted by his alignment parameter. This formulation sets a "survival" level of the personal surplus (embodied in the definition of $z_j^{\bigstar}$) below which even a fully altruistic agent is not willing to sacrifice himself for the benefit of his fellows. Notice that, were it not so, a low benefit agent might deviate and declare high benefit.

For $j = 1$, the agent's objective function is the sum of two components: $U_1\left(\mathbf{t}^1, \mathbf{d}^1\right) + C\left(\mathbf{t}^1, \mathbf{d}^1\right)$, where

$$U_1\left(\mathbf{t}^1, \mathbf{d}^1\right) = z_1^{\bigstar}\left(\mathbf{t}^1, \mathbf{d}^1\right) + t_1 \phi_1\left(\mathbf{t}^1, \mathbf{d}^1\right) \quad (10)$$

and

$$C\left(\mathbf{t}^1, \mathbf{d}^1\right) = \sum_{k=0}^{n} \Pr(k \mid t_1)\sigma(\widehat{k}) E_{t_2^2}\left[z_{s_{\widehat{k}}}^*(\left(t_1, t_2^2\right), d_1^1, \mathbf{d}^{*s_{\widehat{k}}}\left(\mathbf{t}^2, d_1^1\right)) + t_1 \phi_{s_{\widehat{k}}}^*\left(\mathbf{t}^i, d_1^1, \mathbf{d}^{*s_{\widehat{k}}}\left(\mathbf{t}^2, d_1^1\right)\right)\right] + \\ + \sum_{k=0}^{n} \Pr(k \mid t_1)(1 - \sigma(\widehat{k})) E_{t_2^2}\left[z_2^*(\left(t_1, t_2^2\right), d_1^1, \mathbf{d}^{*2}\left(\mathbf{t}^2, d_1^1\right)) + t_1 \Omega_2^*(\left(t_2^2, t_1\right), d_1^1, \mathbf{d}^{*j}\left(\mathbf{t}^2, d_1^1\right), \widehat{k})\right] \quad (11)$$

where $\mathbf{d}^{*j}\left(\mathbf{t}^2, d_1^1\right)$, $j \neq 1$, is the the optimal report of a type $\mathbf{t}^2$ individual who reported d_1^1 in the first period, when the j-group principal plays his equilibrium strategy Ψ^* — that is, $\mathbf{d}^{*j}\left(\mathbf{t}^2, d_1^1\right) = \arg\max_{\mathbf{d}^j} U_j(\mathbf{t}^2, d_1^1, \mathbf{d}^j; \Psi^*)$. The other symbols with an asterisk are similarly defined. Equation (11) defines the expected continuation payoff throughout period 2 and takes into account both the possibility of stabilization and the uncertainty on individual future evaluation of the public good.

Finally, we make the following additional assumptions:

Assumption a1 $h = 1$, $l = 0$

This is an innocuous normalization.

Assumption a2 $c < n - 1$

Only under a2 high type individuals might profit from reporting low type.

Assumption a3 $a = 1$, $e = 0$

To simplify, we consider only the case in which type-a agents assign the same weight to their own material benefit and to their fellows', and type$-e$ agents are indifferent to the fate of their fellows. This allows us to label type-a agents altruists and type-e ones egoists.

Assumption a4 The parameters are such that in an isolated interaction inefficiency obtains at equilibrium.

Asymmetric information is an interesting problem to study only if efficiency cannot be obtained at equilibrium. Below inefficiency is guaranteed at equilibrium by assumption a4$'$ together with assumption a7 (see p. 10 and Appendix A).

Assumption a5 $\left(\sum_{\widehat{v}^j_{\mathbf{d}^j} \in \widehat{\mathbf{v}}^j} \widehat{v}^j_{\mathbf{d}^j} g_j(\widehat{\mathbf{v}}^j, \mathbf{d}^j) - c\right) r_j(\widehat{\mathbf{v}}^j) = 0$

As mentioned, we require ex-post budget balance.

Assumption a6 $t^j_2 - g_j(\mathbf{v}^j, \mathbf{t}^j) \geq 0$, $g_j(\mathbf{v}^j, \mathbf{t}^j) \geq 0$.

This assumption puts a ceiling to the possibility of maneuvering payments for encouraging truth-telling.

3 Analysis

Let us simply state the problem. The principal of a j-group, $j \neq 1$, solves the following program, assuming that truth-telling obtains in the first period.

$$\text{Program } j \; : \quad \max_{\{r_j(\mathbf{v}^2), g_j(\mathbf{t}^2, \mathbf{v}^2); \forall \mathbf{v}^2 \in V, \forall \mathbf{t}^2 \in T\}} \Phi_j \tag{12}$$

s. t.

$$\left(\sum_{\mathbf{t}^2\in\mathbf{T}} v^2_{\mathbf{t}^2} g_j(\mathbf{v}^2,\mathbf{t}^2)-c\right) r_j(\mathbf{v}^2)=0, \qquad \forall\mathbf{v}^2{\in}V \tag{13}$$

$$t^2_2-g_j(\mathbf{v}^2,\mathbf{t}^2)\geq 0, \qquad \forall\mathbf{v}^2{\in}V,\forall\mathbf{t}^2{\in}T \tag{14}$$

$$0\leq g_j(\mathbf{v}^2,\mathbf{t}^2) \qquad \forall\mathbf{v}^2{\in}V,\ \forall\mathbf{t}^2{\in}T \tag{15}$$

$$0\leq r_j(\mathbf{v}^2)\leq 1, \qquad \forall\mathbf{v}^2\in V \tag{16}$$

$$U_j\left(\mathbf{t}^2,\mathbf{t}^2\right)\geq 0, \qquad \forall\mathbf{t}^2{\in}T \tag{17}$$

$$U_j\left(\mathbf{t}^2,\mathbf{t}^2\right)\geq U_j\left(\mathbf{t}^2,\mathbf{d}^j\right), \qquad \forall(\mathbf{t}^2,\mathbf{d}^j)\in T\times T \tag{18}$$

Conditions (13)-(16) simply requires feasibility within our framework. Inequality (17) imposes individual rationality and inequality (18) requires that truth-telling is incentive compatible.

Consider now the principal's problem at $j=1$ (in the following an asterisk denotes an object evaluated at the equilibrium values of program j).

The first-period principal solves the following program, assuming that the other first-period principals, if any, are playing the strategy $\overline{\boldsymbol{\sigma}}\,(k)\,,k=0,...,n$:

$$\text{Program 1}\ :\ \max_{\{r_1(\mathbf{v}^1),g_1(\mathbf{v}^1,\mathbf{t}^1),\forall\mathbf{v}^1\in V,\forall\mathbf{t}^1\in T;\boldsymbol{\sigma}(k),k=0,\dots,n\}}\Phi_1+$$

$$+\sum_{k=0}^{n}\Pr(k)\left\{\sigma(k)\Phi^*_{s_k}+(1-\sigma(k))E_{t^2_2}\left[kz^*_2\left(\left(a,t^2_2\right),a,\left(a,t^2_2\right)\right)+(n-k)z^*_2(\left(e,t^2_2\right),e,\left(e,t^2_2\right))\right]\right\}$$

s. t.

$$\left(\sum_{\mathbf{t}^1\in\mathbf{T}} v_{\mathbf{t}^1} g_1(\mathbf{v}^1,\mathbf{t}^1)-c\right) r_1(\mathbf{v}^1)=0,\forall\mathbf{v}^1{\in}V \tag{19}$$

$$t^1_2-g_1(\mathbf{v}^1,\mathbf{t}^1)\geq 0, \qquad \forall\mathbf{v}^1{\in}V,\forall\mathbf{t}^1{\in}T \tag{20}$$

$$0\leq g_1(\mathbf{v}^1,\mathbf{t}^1), \qquad \forall\mathbf{v}^1{\in}V,\forall\mathbf{t}^1{\in}T \tag{21}$$

$$0\leq r_1(\mathbf{v}^1)\leq 1, \qquad \forall\mathbf{v}^1{\in}V \tag{22}$$

$$U_1\left(\mathbf{t}^1,\mathbf{t}^1\right)+C\left(\mathbf{t}^1,\mathbf{t}^1\right)\geq U_1\left(\mathbf{t}^1,\mathbf{d}^1\right)+C\left(\mathbf{t}^1,\mathbf{d}^1\right), \qquad \forall(\mathbf{t}^1,\mathbf{d}^1)\in T\times T \tag{23}$$

Similar to program j definition, conditions (19)-(22) impose feasibility and inequality (23) is the intertemporal incentive compatibility constraint. Of course, at a symmetric Nash equilibrium it will be $\boldsymbol{\sigma}\,(k)=\overline{\boldsymbol{\sigma}}\,(k)\,,k=0,...,n$.

To simplify the analysis we make a further assumptions.

Assumption a7 Groups size is $n = 2$.

As a result, the number of possible type profiles (states) is greatly reduced. We will comment on the generalization to the case of groups of any size in section 4.

When $n = 2$, assumption a4 simplifies to:

Assumption a4' The probability p that an agent is of type h is such that $p > \frac{2-2c}{2-c(1+\theta)}$.

Now we have all the tools to develop our analysis. The first question we study is whether a first-period principal may find it profitable to stabilize the group and communicate his information to the second-period principal. Preliminarily, let us state a simple fact.

Fact 1 *In an isolated interaction, the group expected surplus is larger when the permanent type of each member is common knowledge than when it is private information.*

Proof. See Appendix, part A. ■

It follows from Fact 1 that if eliciting information on permanent types were not costly, then group stabilization would certainly benefit society, because in the second period common knowledge would induce higher surplus. However, notice that from a first-period principal's perspective, when a group is composed of two type-e agents group surplus would be larger when the group is dissolved and members go to fresh groups, but this would be detrimental for society. Therefore the values of information for a group and for the whole society do not coincide.

Nevertheless, knowledge of group composition would be an asset that one would like to exploit, choosing whether stabilizing or dissolving the group. In a context of asymmetric information, extracting private information from the agents is costly. Does this cost exceed the benefits of common knowledge? We shall provide an answer in the next subsections.

3.1 Only one active group.

The analysis of the whole game is rather complex, so it is illuminating to study first the problem under the hypothesis that only one group exists.

We summarize our findings in the following.

Proposition 1 *When only one group exists, the knowledge of group composition acquired through the first interaction cannot be exploited to increase per-period surplus with respect to the case of an isolated interaction. Formally, the first period principal never stabilizes the group, i.e.* $\sigma(k) = 0,\ k = 0, .., 2$. *Such result is due to the fact that when group stabilization is possible, the* *cost of inducing truthful reporting exceeds the efficiency gain.*

The same holds when many groups are supervised by a unique principal who aims at maximizing the surplus of the society.

Proof. The equilibrium allocations in stabilized groups coincide with those we characterized in proving Fact 1 under common knowledge of the individuals' permanent type (see Appendix A). As fresh groups are concerned, the second-period principal computes the probability, Θ, that a member is of type-a by means of the Bayes' rule, under the assumption that at date 1 all agents reported their type truthfully, that is $\Theta = \frac{\theta^2(1-\sigma(2))+2\theta(1-\theta)(1/2)(1-\sigma(1))}{\theta^2(1-\sigma(2))+2\theta(1-\theta)(1-\sigma(1))+(1-\theta)^2(1-\sigma(0))}$. Now, suppose $\sigma(k) = 0$, $k = 0,..,2$ at equilibrium. Then, it follows that $\Theta = \theta$, and hence the equilibrium allocation coincides with that we found in proving Fact 1 under asymmetric information.

In order to prove that at equilibrium it is really $\sigma(k) = 0, k = 0,..,2$, we show that the optimal strategies of the first-period principal belong to the subset of strategies such that no communication to the second-period principal will be undertaken.

The proof is not immediate. Indeed, the trick used in the proof of the ratchet effect in Fudenberg and Tirole,1991, of constructing an equivalent static strategy through an appropriate averaging of the candidate optimal dynamic strategy, does not apply in our game. Precisely, no averaging can simultaneously conserve both the surplus of the group and that of each agent at their initial levels. So, we proceed in two steps. In the first step we consider a modified mechanism in which the mentioned stratagem applies. In the second step we show that the optimal solution of the modified mechanism also maximizes the original mechanism.

Step 1

Consider a new mechanism, which differs from ours only in the timing of reporting: firstly each agent confidentially reports his permanent type to the principal, then each of them learns his benefit from the public good and reports it without knowing the permanent type of his mates — except, of course, in stabilized groups in which no reporting of the permanent type is needed and members' permanent reported types are common knowledge. Let $\Psi^e = \left\{r^e_j(\widehat{\mathbf{v}}^j), g^e_j(\widehat{\mathbf{v}}^j, \mathbf{d}^j), j = 1, 2, s_0, s_1, s_2, \sigma(k), k = 0, 1, 2, \text{all } \widehat{\mathbf{v}}^j, \text{all } \mathbf{d}^j\right\}$ denote an equilibrium strategy in this modified game , and let lotteries $\gamma_j(\mathbf{t}^i, \mathbf{d}^j, \widehat{\mathbf{v}}^j) = \left\{r^e_j(\widehat{\mathbf{v}}^j), 1 - r^e_j(\widehat{\mathbf{v}}^j); t^i_2 - g^e_j(\widehat{\mathbf{v}}^j, \mathbf{d}^j), 0\right\}$ denote the surplus allocation for a j-group to an individual of type $\mathbf{t}^i$, $i = 1, 2$, who declares $\mathbf{d}^j$, when the reported state is $\widehat{\mathbf{v}}^j$, $j = 1, 2, s_0, s_1, s_2$.

The equilibrium surplus allocation in the whole game that a type-e individual expects before learning his private benefit is

$$
\begin{aligned}
S\left(e;e,\left\{\gamma_j\right\}\right) = p\theta\left[p\gamma_1(eh,eh,eh.ah) + (1-p)\,\gamma_1(eh,eh,eh.al)\right] + \\
+\,\theta p^2\left[\sigma(1)\gamma_{s1}(eh,eh,eh.ah) + (1-\sigma(1))\,\gamma_2(eh,eh,eh.ah)\right] + \\
+\,\theta p(1-p)\left[\sigma(1)\gamma_{s1}(eh,eh,eh.al) + (1-\sigma(1))\,\gamma_2(eh,eh,eh.al)\right] + \\
+\,(1-\theta)\,p\left[p\gamma_1(eh,eh,2eh) + (1-p)\,\gamma_1(eh,eh,eh.el)\right] + \\
+\,(1-\theta)\,p^2\left[\sigma(0)\gamma_{s0}(eh,eh,2eh) + (1-\sigma(0))\,\gamma_2(eh,eh,2eh)\right] + \\
+\,(1-\theta)\,p(1-p)\left[\sigma(0)\gamma_{s0}(eh,eh,eh.el) + (1-\sigma(1))\,\gamma_2(eh,eh,eh.el)\right] \quad (24)
\end{aligned}
$$

Now let us consider the class of strategies for the first-period principal such that $\sigma(k) = 0$, $k = 0,..,2$. Observe that if a strategy in this class results to be optimal, also the second-period principal will be using it at equilibrium — we are assuming uniqueness for simplicity; of course, choosing a value for $\sigma\left(k\right)' s$ is inessential in the second period. We name a strategy with $\sigma(k) = 0$, $k = 0,..,2$, a static strategy.

We will show that, whatever the equilibrium allocation set $\left\{\gamma_j\right\}$,we can always attain the principal's problem value by restricting attention to static strategies. This means that using information extracted in the first period is worthless.

Let $\gamma(\mathbf{t}^1, \mathbf{d}^1, \widehat{\mathbf{v}}^1)$ be the allocation, induced by an optimal strategy in the class of static strategies, to an individual of type $\mathbf{t}^1$ who reports $\mathbf{d}^1$ in state $\widehat{\mathbf{v}}^1$.

Next, consider an individual who is of type $\mathbf{t}$ in the first period and rewrite his surplus allocation for the whole game, let it be denoted by $S\left(t_1, t_1; \{\gamma\}\right)$, when $\gamma()$ and not $\gamma_j()$, for all j, is adopted. Notice that the expected value of $S\left(t_1, t_1; \{\gamma\}\right)$ (expectation is computed with respect to the probabilities of provision) equals $E_{t_2^1}\left[U_1\left(\mathbf{t}^1, \mathbf{t}^1\right) + C\left(\mathbf{t}^1, \mathbf{t}^1\right)\right]$. As an example, the equilibrium expected surplus of a type e individual is

$$
\begin{aligned}
S\left(e,e;\{\gamma\}\right) = p\theta\left[p\gamma(eh,eh,eh.ah) + (1-p)\,\gamma(eh,eh,eh.al)\right] + \\
+\,\theta p^2\gamma(eh,eh,eh.ah) + \theta p(1-p)\gamma(eh,eh,eh.al) + \\
+\,(1-\theta)\,p\left[p\gamma(eh,eh,2eh) + (1-p)\,\gamma(eh,eh,eh.el)\right] + \\
+\,(1-\theta)\,p^2\gamma(eh,eh,2eh) + (1-\theta)\,p(1-p)\gamma(eh,eh,eh.el) \quad (25)
\end{aligned}
$$

Notice that $S\left(e,e;\{\gamma\}\right)$ can be obtained from $S\left(e,e;\left\{\gamma_j\right\}\right)$ by writing $\gamma()$ instead of γ_j in all states, and then simplifying.

It will be $S\left(e,e;\{\gamma\}\right) = S\left(e,e;\left\{\gamma_j\right\}\right)$ if the following set of elementary conditions on $\{\gamma\}$ is satisfied:

$$\gamma(eh, eh, eh.\boldsymbol{\tau}) = \frac{\gamma_1(eh, eh, eh.\boldsymbol{\tau})}{2} + \\ + \frac{1}{2}\sum_{k=0}^{2}\Pr(k \mid e.\boldsymbol{\tau})\left\{\sigma\left(k\right)\gamma_{s_k}\left(eh, eh, eh.\boldsymbol{\tau}\right) + \left(1-\sigma\left(k\right)\right)\gamma_2\left(eh, eh, eh.\boldsymbol{\tau}\right)\right\} \quad \forall\boldsymbol{\tau}. \tag{26}$$

In the same way we get:

$$\gamma(ah, ah, 2ah) = \frac{\gamma_1(ah, ah, 2ah)}{2} + \\ + \frac{1}{2}\left[\sigma\left(2\right)\gamma_{s2}(ah, ah, 2ah) + \left(1-\sigma\left(2\right)\right)\gamma_2(ah, ah, 2ah)\right] \tag{27}$$

$$\gamma(ah, ah, ah.eh) = \frac{\gamma_1(ah, ah, ah.eh)}{2} + \\ + \frac{1}{2}\left[\sigma\left(1\right)\gamma_{s1}(ah, ah, ah.eh) + \left(1-\sigma\left(1\right)\right)\gamma_2(ah, ah, ah.eh)\right] \tag{28}$$

$$\gamma(ah, ah, ah.el) = \frac{\gamma_1(ah, ah, ah.el)}{2} + \\ + \frac{1}{2}\left[\sigma\left(1\right)\gamma_{s1}(ah, ah, ah.el) + \left(1-\sigma\left(1\right)\right)\gamma_2(ah, ah, ah.el)\right] \tag{29}$$

$$\gamma(ah, ah, ah.al) = \frac{\gamma_1(ah, ah, ah.al)}{2} + \\ + \frac{1}{2}\left[\sigma\left(2\right)\gamma_{s2}(ah, ah, ah.al) + \left(1-\sigma\left(2\right)\right)\gamma_2(ah, ah, ah.al)\right] \tag{30}$$

$$\gamma(\cdot l, \cdot l, \cdot l.\boldsymbol{\tau}) = \{\cdot, \cdot; 0, 0\} \tag{31}$$

These conditions are such that the strategy that implements this $\{\gamma\}$ is well defined: that is, both its $r(\mathbf{v})$'s and $g(\mathbf{v},\mathbf{t})$'s are defined in terms of all the equilibrium probabilities $r_j(\mathbf{v})$ and all the payments $g_j(\mathbf{v},\mathbf{t})$ in such a way that the allocation for a given type in a given state uses the same probabilities of providing the good as the other types' allocations in the same state. Moreover each of these allocations is feasible, because it is a convex linear combination of feasible allocations contingent on the same state – so all constraints on payments and ex-post budget balance constraints are satisfied.

Now let us prove that the strategy that implements $\{\gamma\}$ is incentive compatible. Unilateral deviations from truthful reporting to be considered are (indicating the true type on the left side and the false type on the right side): $e \to a, a \to e, eh \to el, ah \to al$.

Since by construction, each type's expected payoff from truthful reporting is the same under $\{\gamma\}$ and $\{\gamma_j\}$, the strategy that implements $\{\gamma\}$ is incentive compatible if the expected value of lottery $S(t_1,d_1;\{\gamma\})$ is lower than the expected value of lottery $S(t_1,d_1;\{\gamma_j\})$ for each possible t_1,d_1. Now, notice that the strategy that implements $\{\gamma\}$ must induce an individual of type $\mathbf{t}^1$ at date 1 to report $\mathbf{t}^1$. Instead the strategy that implements $\{\gamma_j\}$, $j=1,2,s_{k,}$ $k=0,1,2$ must induce an individual that is of type $\mathbf{t}^1$ at date 1 not only to report his true type in group 1 but also to report his true types in all the possible second period groups. Therefore the strategy which implements $\{\gamma\}$ offers less opportunities for lying that the strategy that implements $\{\gamma_j\}$. Thus we conclude that the expected value of lottery $S(t_1,d_1;\{\gamma\})$ is lower than the expected value of lottery $S(t_1,d_1;\{\gamma_j\})$.

Finally, noticing that the first-period principal gets the same group surplus as with $\{\gamma_j\}$, we conclude that $(\{\gamma\},\{\gamma\})$ is not worse than $\{\gamma_j\}$, $j=1,...,2$. Actually, the first-period principal's surplus coincides with the altruist expected payoff if at least one altruist belongs to the group and with twice the egoist expected payoff if only egoist agents compose the group. Given that our construction conserves total individual payoff for all types in any state, it also conserves first-period principal surplus[4].

Step 2.

Let us come back to our original game, in which each agent learns and reports both dimensions of his type simultaneously. Note that the incentive constraints in the first-period mechanism are more restrictive than those considered in the modified mechanism of step 1, since the deviations to be considered are $eh \to al, eh \to el, eh \to ah, el \to al, ah \to eh, ah \to al, ah \to el, al \to el$. Notice that our assumption that material surplus must be nonnegative implies that no l type will ever report h type.

[4]Recall that a high benefit altruist's material surplus is one half of his surplus.

It follows that in the class of the strategies that use the information elicited in the first period no strategy can exist that does better than the best strategy in the modified game above.

Let us consider the class of the strategies that do not use the information elicited in the first period. We will show that the system of incentive compatibility constraints in our game is equivalent to that in the modified game. In what follows we directly replace in the ICs the optimal payments, which are described in Appendix A. Note that:

1) Incentive compatibility constraint $IC(e,a)$ in the modified game is equivalent to the two constraints $IC(eh,al)$, $IC(eh,ah)$ in the original game, since a low-type egoist does not have any incentive to lie. In fact in both games $r(\cdot h,\cdot h)=1$, $r(\cdot l,\cdot l)=0$ at equilibrium and hence $IC(e,a)$ is

$$
\begin{aligned}
&p\theta(p+(1-p)(1-c)r(eh.al))+p\,(1-\theta)\,(p(1-\frac{c}{2})+(1-p)(1-c)r(eh.el))\geq\\
&\quad p\max\Big\{\theta p(1-\frac{c}{2})+(1-p)(1-c)r(ah.al))+(1-\theta)\,(p(1-c)+(1-p)(1-c)r(ah.el)),\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\theta pr(ah.al))+(1-\theta)pr(eh.al)\} \qquad (32)
\end{aligned}
$$

The preceding IC reads as follows. Take an egoist (addressed also as first player). The left-hand-side represents the payoff he obtains by truthfully reporting his permanent type. Suppose he is characterized by high benefit (an event with probability p). He will obtain a positive surplus, the amount of which depends on the bi-dimensional type of his fellow. Conversely, an egoist with low benefit (an event with probability $1-p$) will obtain no surplus independently of his fellow's type. Let us discuss now how the positive expected surplus is formed. If the fellow has type (a,h), whose probability is θp, the public good will be certainly provided. The egoist under scrutiny will get 1 at no cost. All the cost of provision is paid by his fellow. When the fellow has type (a,l) the good is provided with probability $r(eh.al)$, but now the full cost is paid by the first player. Moreover, also the fellow can be of type-e. With probability p he will be of (e,h) type. In this case the good is certainly provided and the cost is equally shared. With probability $1-p$ the fellow is of (e,l) type: now the good is provided with probability $r(eh.el)$ and all the cost is entirely born by the first player.

The right-hand-side represents agent's surplus when he lies on his permanent type. Once more he gets a positive surplus only when his own contingent type is h. Note that he has two possible lies: telling to be of (a,h) type or of (a,l) type. He will clearly opt for the one that secures more surplus to himself. Consider the possible cases. First, the egoist declares (a,h). When his fellow is of (a,h) type, the public good will be certainly provided and the cost is equally shared, because both individuals appear to have the same type. When the fellow is of (a, l) type the good is provided with probability $r(ah.al)$ and all the cost is paid by the first player. Again, when the fellow is of (e, h) type the good is certainly provided and the cost is paid by the first player because he declared to be altruist. Finally, when the fellow is of (e, l) type, the good is provided with probability $r(ah.el)$ and the cost is born by the first player. Suppose now the egoist declares to be of (a, l) type. If the fellow is of (a, h) type the good is provided with probability $r(ah.al)$ at the fellow's expenses. When the fellow is of (e, h) type the good is provided with probability $r(eh.al)$ again at the fellow's expenses. Finally, note that when the fellow has contingent type l the good will be certainly not provided.

This incentive constraint is equivalent to the system of constraints $IC(eh, al)$ and $IC(eh, ah)$, that is

$$\theta(p + (1-p)(1-c)r(eh.al))+ \\ + (1-\theta)\left(p(1-\frac{c}{2}) + (1-p)(1-c)r(eh.el)\right) \geq \theta pr(ah.al)) + (1-\theta)pr(eh.al) \quad (33)$$

$$\theta(p + (1-p)(1-c)r(eh.al)) + (1-\theta)\left(p(1-\frac{c}{2}) + (1-p)(1-c)r(eh.el)\right) \geq \\ \theta(p(1-\frac{c}{2}) + (1-p)(1-c)r(ah.al)) + (1-\theta)\left(p(1-c) + (1-p)(1-c)r(ah.el)\right) \quad (34)$$

2) The system of incentive compatibility constraints $IC(a, e)$ and $IC(eh, el)$ in the modified game is equivalent to the system of constraints $IC(eh, el)$ and $IC(ah, eh)$ in the original game. In fact since in both games $r(\cdot h, \cdot h) = 1, r(\cdot l, \cdot l) = 0$ at equilibrium, and a high altruist will certainly report h in both games, $IC(a, e)$ is

$$(1-c)(1-p)p(\theta r(ah.al) + (1-\theta)r(eh.al) - \theta r(ah.el) - (1-\theta)r(eh.el))+ \\ + p\theta\left(p(2-c)(1-1) + (1-p)(1-c)\left(r(ah.al) - r(eh.al)\right)\right)+ \\ p(1-\theta)(p(2-c)(1-1) + (1-p)(1-c)(r(ah.el) - r(eh.el))) \geq 0 \quad (35)$$

which simplifies to

$$\theta\left(r(ah.al) - r(ah.el)\right) + (1-\theta)\left(r(eh.al) - r(eh.el)\right) + \\ + \theta(r(ah.al) - r(eh.al)) + (1-\theta)(r(ah.el) - r(eh.el)) \geq 0 \quad (36)$$

On the other hand $IC(eh, el)$ is

$$\theta r(eh.al) + (1-\theta) r(eh.al) - \theta r(ah.el) - (1-\theta)\, r(eh.el) \geq 0 \tag{37}$$

and $IC(ah, eh)$ is

$$\theta r(ah.al) + (1-\theta)\, r(ah.el) - \theta r(eh.al) - (1-\theta)\, r(eh.el) \geq 0 \tag{38}$$

and hence adding these two constraints we get

$$\begin{aligned} \theta r(ah.al) + (1-\theta) r(eh.al) - \theta r(ah.el) - (1-\theta)\, r(eh.el) + \theta r(ah.al) + \\ + (1-\theta)\, r(ah.el) - \theta r(eh.al) - (1-\theta)\, r(eh.el) = \\ \theta\, (r(ah.al) - r(ah.el)) + (1-\theta)\, (r(eh.al) - r(eh.el)) + \\ + \theta (r(ah.al) - r(eh.al)) + (1-\theta)(r(ah.el) - r(eh.el)) \geq 0 \end{aligned} \tag{39}$$

3) $IC(al, el)$, $IC(el, al)$ and $IC(eh, ah)$ are not binding at equilibrium. (see Appendix part A) Therefore we have proved that the system of IC constraints in the modified mechanism is equivalent to that of the original mechanism. This allows us to conclude that the class of the strategies for the modified mechanism that do not use the information elicited in the first period is large enough to include the optimal strategy for our original game. ■

A careful examination of the proof above shows that dissolving the group is an equilibrium strategy no matter what the number of altruists and the number of existing groups are, provided that the equilibrium strategy used in fresh groups and in first-period groups is the same. It is immediately seen that the proof logic also applies to the case where one first-period principal supervises all first-period groups with the purpose of maximizing the overall surplus.

The result just obtained is due to a complex sort of ratchet effect. In the standard framework, the ratchet effect consists in the agent anticipating that in next rounds of interaction the principal will exploit the information elicited in the current round, so there is an incentive for the agent to lie on his type; restoring the incentive to telling the truth entails an additional cost to the principal that overcomes the benefit of exploiting the information. In our framework the story is more complicated, since the expectation that the principal will intertemporally exploit information triggers three distinct perverse incentive effects.

The first - that we call "plain ratchet effect" - corresponds to the situation described above: good-type agents are reluctant to declare truthfully their permanent type fearing that the principal will use it to their disadvantage in the current and in the following rounds. However setting $a = 1$ and $e = 0$ as we did for simplicity, causes this effect to vanish; in fact, perfectly altruist agents do not care about possible reductions of their own surplus, as they look at the whole group surplus. This allows us to focus on the remaining two effects, that are much less straightforward.

The second effect - that we call "chain of lies effect" consists in an agent misreporting his permanent type in the first period in order to get the chance of advantageously lying on his temporary type in the second period. This effect only regards an egoist: he can falsely declare himself to be altruist in the first period to increase the probability of group stabilization; if the group is indeed stabilized, he will report low benefit in case he discovers himself to be a high benefit individual. Specifically, as a consequence of the first lie he will face an incentive constraint designed for an altruist, and therefore inappropriate for inducing an egoist to tell the truth. An obvious case in which this effect does prevent the exploitation of the information elicited during the first period, arises when the principal sets $\sigma(k) = 1$, for all k - recall that Fact 1 ensures that in this case the group expected surplus in the second period is greater than in the case $\sigma(k) = 0$ for all k. For simplicity, let us consider the case of a single group. An altruist has no reason to lie: in fact, by telling the truth he can only end up in groups s_1 and s_2 and in both such groups efficiency results; on the opposite, by lying he risks to be part of group s_0 where inefficiency prevails. Instead, egoists find it profitable to lie. This is an immediate implication of Proposition 1 — still, we offer an independent proof in the Appendix B. Crucially, note that it is the possibility of a second lie that makes the incentive constraint of egoists especially hard to satisfy. In fact, once the egoist has declared altruist, he profits from declaring low even when he is high[5], so the perspective of the second lie increases the attractiveness of the first[6].

[5] Once in a s_1-group a false altruist who happens to be high obtains p by lying and $1 - c$ by telling the truth. We know that $p > 1 - c$, as this inequality is implied by the inefficiency condition, that can be rewritten as $p > (1 - c)\frac{2}{2-(1+\theta)c}$.

In a s_2-group the corrisponding condition is $p - [p(1 - c/2) + (1 - p)(1 - c)] = -\frac{1}{2}pc - 1 + c + p > 0$, which is satisfied for $p > \frac{2-2c}{2-c}$. This inequality also follows from the inefficiency condition, since $\frac{2-2c}{2-(1+\theta)c} > \frac{2-2c}{2-c}$.

[6] For some triplets of parameters, absent the possibility of the second lie, the egoist would not find it advantageous to tell the first lie in front of the announcement $\sigma_0 = \sigma_1 = \sigma_2 = 1$. In fact the payoff of truthtelling is $p\left[\theta\left(p + (1 - p)(1 - c)\right) + (1 - \theta)\left[p\left(1 - c/2\right) + (1 - p)(1 - c)\frac{1}{2}p\frac{-2+c}{1-c-2p+pc}\right]\right]$, while that of the single lie would be $p\left[\theta p(1 - c/2) + \theta(1 - p)(1 - c) + (1 - \theta)(1 - c)\right]$. In fact, for $p = 3/4, c = 3/4, \theta = 3/4$ the two expressions amount to 0.3457 and 0.5529, respectively.

The third effect - probably the most surprising at first sight - concerns altruists: even a perfect altruist, who does aim at maximizing group surplus, may be tempted to declare to be an egoist to modify the probabilities of public good provision, in order to enforce what she myopically sees as an improvement in group welfare. We label this effect "white lie effect". The point is that the principal is assumed to have a commitment to implement - in a time consistent way - the strategy initially announced; agents, instead, are by definition followers, who maximize ex-post, taking the principal's announcement as given. So, in her reporting decision an agent whose objective function is perfectly aligned with the principal's, does not realize - qua agent - that the latter is subjecting himself to a costly constraint just for disciplining the egoists, and tries to improve upon the resulting outcome. The consequence is the opposite, since, in order to prevent such behavior, the principal is forced to subject himself to further constraints, and therefore to shift to a less preferred strategy.

Consider the following example, in which one principal controls all N first-period groups and N is infinitely large - so the change in the expected composition of fresh groups due to an agent's deviation is negligible - and maximizes the overall surplus. Let $c = p = \theta = 3/4$. When $\sigma(k) = 0$, for all k, efficiency can never be obtained[7]. However, setting $\sigma(0) = 0$, $\sigma(1) = 1$, $\sigma(2) = 0$ would ensure full efficiency in the second period. In fact, not surprisingly, efficiency is warranted in group s_1 and, furthermore, thanks to the fact that the proposed selective group stabilization leads to a Θ sufficiently greater than θ, efficiency would also obtain in a fresh group[8]. It can easily be shown that this move does not deteriorate the incentive for an egoist to declare truthfully her permanent type in the first period - indeed, an egoist who happens to be high and has declared her permanent type truthfully, in the second period obtains nearly the maximum conceivable surplus[9]. Instead, it is just the altruists who now have a clear incentive to lie: whoever an altruist's mate, group surplus is greater if the altruist declares egoist. To better see the point, let us consider the two possible cases. First case: the altruist has an altruist mate. The group will be stabilized thanks to the former altruist's lie, and thus he obtains the efficient surplus.

[7]The efficiency condition $p < \frac{2-2c}{2-c(1+\theta)}$ is not satisfied, as $.75 > .727\,27$.

[8]Since N is large, $\Theta = \frac{\theta^2}{\theta^2+(1-\theta)^2} = 0.9$ and therefore the threshold for p would become $p < \frac{2-2c}{2-c(1+\Theta)} = 0.869\,57$, so the efficiency condition would be satisfied – recall that it is $p = .75$.

[9]More precisely - remember that her payoff is only nonzero when she happens to be high - it can be shown that by lying she has only to lose. In fact, if her mate is an altruist, she would obtain the payoff of an egoist in the fresh group, rather that the payoff of an egoist in a s1 group, which is greater (there she is sure of being matched with an altruist); if instead her mate is an egoist, the group would be stabilised, so the best payoff she can get is by declaring low (again the chain of lies) and is equal to p (in fact by declaring high, as a supposed altruist, she would be asked to pay all the cost of the good, and thus obtain $1 - c$, which is less than p since the assumption $p > 2\frac{1-c}{2-c}$ implies $p > 1 - c$) while by declaring truthfully, she would obtain the payoff of an egoist in the fresh mechanism, which in this case is greater (indeed for $p = c = 3/4$ and $\Theta = 9/10$ it can easily be checked that it is $\Theta p + (1 - \Theta)\, p\,(1 - c/2) + (1 - p)(1 - c) > p$).

Instead, by telling the truth, both fellows would be reallocated to fresh groups, where they would get each less than half of the efficient pie. Second case: the altruist has an egoist mate. The lie will cause reallocation to fresh groups, which is preferable to stabilization (stabilization would only ensure the efficient group surplus, while reallocation in fresh groups allows one altruist and one egoist to obtain more because efficiency is also secured and the ratio of altruists to egoists exceeds 1:1).

3.2 Infinitely many active groups.

We now turn to the case in which N groups exist and, according to the initial description of the game, each first-period principal aims at maximizing the surplus of the group he supervises. Since his first-period choices concur in determining the composition of second-period fresh groups, a strategic interaction between first-period principals arises. In this context Proposition 1 cannot be assumed to hold, since no game among first-period principals is considered there. So, group stabilization cannot be excluded. Of course, Proposition 1 implies that stabilization can only deteriorate social surplus.

In the following, the number of groups is so large that each first-period principal can neglect the impact of his own stabilization decisions on the composition of fresh groups.

Before presenting our results, let us explain, on intuitive grounds, why the strategy of never stabilizing groups may not be Nash. If all the other first-period principals dissolved their groups, one principal might find it advantageous to selectively stabilize his group: in particular, when both members are of type-e, letting them go is attractive, since they will gain surplus in fresh groups at the expense of type-a agents coming from other groups; conversely, stabilizing the group when this is composed of two type-a individuals ensures more surplus than reallocation into fresh groups. So, a free riding problem characterizes the interaction between first-period principals. Nevertheless, group stabilization does not necessarily occur in equilibrium, since eliciting and exploiting private information on members' permanent type is still made costly by the ratchet effect.

As regards the computation of Nash equilibria, thanks to N being infinitely large we can easily verify whether a given strategy for first-period principals is Nash by solving a simple linear programming problem — as we saw, a first period principal knows that the composition of fresh groups is independent of his own decision. Below we present an example in which group stabilization obtains at equilibrium, and an example in which stabilization does not obtain.

Let be $\theta = 1/2$, $c = 3/4$, $p = 3/4$, $n = 2$, N infinitely large. The strategy of always dissolving the group is not Nash since the best reply of a first-period principal to such a strategy is $\sigma(0)=0$, $\sigma(1)=1$, $\sigma(2)=\frac{11}{15}$, which would get him an expected surplus of 1.610 — notice that he will set $r_1(eh.al)=.7416$, $r_1(ah.al)=.9841$, $r_1(eh.el)=.4674$, $r_1(ah.el)=1$, $r_1(\cdot h.\cdot h)=1$, $r_1(\cdot l.\cdot l)=0$; instead the strategy of always dissolving the group brings him $2\,(786222)=1.572$ only. Therefore unilateral deviation is profitable.

Then, consider stabilizing the group with probability $\sigma(0)=0$, $\sigma(1)=1$, $\sigma(2)=.4720$. A first-period principal knows that, independently of his move, the probability that a member of a fresh group is of type a is $\Theta=\frac{\theta^2(1-\sigma(2))+2\theta(1-\theta)(1/2)(1-\sigma(1))}{\theta^2(1-\sigma(2))+(1-\sigma(1))2\theta(1-\theta)+(1-\sigma(0))(1-\theta)^2}$, that is $\Theta=.34627$. If we compute a first-period principal's optimal reply we find $\sigma(0)=0$, $\sigma(1)=1$, $\sigma(2)=.4720$, and so a Nash equilibrium occurs – he will set $r_1(ah.al)=1$, $r_1(eh.el)=.1386$, $r_1(eh.al)=.6944$, $r_1(ah.el)=1$. Group expected surplus at equilibrium is $0.7799+.7803=1.5602$. Notice that, as we expected, group expected surplus at equilibrium is less than it would be if always dissolving the group could be an equilibrium strategy.

Instead consider the parameter constellation $\theta=\frac{1}{4}$, $p=\frac{3}{4}$, $c=\frac{3}{4}$. The strategy of stabilizing the group with probability $\sigma(0)=0$, $\sigma(1)=0$, $\sigma(2)=0$ is Nash when N is infinitely large. Indeed a first-period principal's optimal reply is $\sigma(2)=0$, $\sigma(1)=0$, $\sigma(0)=0$[10].

We summarize our results in the following:

Proposition 2 *When a very large number of groups exist, group stabilization can be obtained in a Nash equilibrium. Therefore the knowledge of group composition acquired in the first interaction has a positive economic value for the group.*
However, per period expected group surplus cannot be greater than in a single isolated interaction.

4 Discussion and extensions

First of all, notice that removing the assumption that the levels of benefit, which type-h and type-l players derive from the public good, are the same across periods, does not invalidate Proposition 1. Moreover Proposition 1 still holds if a discount factor different from 1 is introduced, and also if a greater (eventually infinite) number of periods is considered. Notice, in this regards, that no stable groups can dissolve after the second interaction since the share of altruists in fresh groups may only decreases over time. Also the assumption that group size is $n=2$ can be easily generalized. We conjecture that at Nash equilibrium only groups in which the share of altruists

[10] The probability that a member is of type a in a fresh group is $\Theta=\frac{1}{4}$. The probabilities of provisionn are $r_i(\cdot h.\cdot h)=1$, $r_i(\cdot l.\cdot l)=0$, $r_i(eh.al)=.712$, $r_i(ah.el)=.712$, $r_i(eh.el)=.808$, $r_i(ah.al)=1$ $i=1,2$.

is between a lower and an upper bound might be stabilized.

More interesting is to reconsider the assumptions concerning: a) the principal's powers; b) the definition of altruism

4.1 Principal's powers.

In principle, we may imagine several ways to modify principal's powers. Below, we content ourselves to examine one situation in which the principal has more powers and one in which he has less powers than we have assumed until now.

First: more powers. Recall the assumption that first-period principals do not set the second-period groups allocations. Furthermore second-period principals are assumed to hold no more information about permanent types than group members.

The first assumption is made only for simplicity and is not essential for the results above. In fact, assuming mechanism coordination across periods can improve the value of the best dynamic strategy, but Proposition 1 continues to apply (the proof above holds whatever the candidate optimum strategy). On the other hand, when first period principals play Nash, giving them greater powers in case of group stabilization could only enlarge the set of parameter triplets for which in the symmetric equilibrium group stabilization obtains.

The second assumption rules out the possibility of an informed second-period principal who might prefer not to reveal information to agents. Further research is needed to analyze this possibility.

Second: less powers. The principal might not be endowed with the power of stabilizing groups, since this decision might be thought to be in the hands of group members. Of course, this requires that group composition is common knowledge after the first period. The effects of this alternative power distribution are easily understood in our framework. After the first round agents will unanimously choose to terminate their relationship when no altruist is present, and to continue it when there are either one or two altruists[11]. So - irrespectively of whether one principal controls symmetrically all groups, or instead N principals play Nash - in equilibrium no altruist will be present in fresh groups.

[11]It is immediate to intuit why two altruists prefer to stay together (so they split the efficient surplus, instead of receiving each less than half of it in fresh groups) and two egoists have nothing to gain from staying together (that is the worst group composition from their point of view). When only one of them is an altruist, the egoist prefers continuation, since it ensures him the best possible payoff. As to the altruist, he must compare the efficient group surplus secured by the s_1-group with the sum of the expected material payoff of one altruist and one egoist in a fresh group. It is clear that this sum is less than the efficient group surplus as soon as $\Theta < 1/2$ (in that case one altruist and one egoist together receive less than expected group surplus, which is at most the efficient surplus). The last step is to observe that when $\sigma(2) = 1$ and $\sigma(0) = 0$ it is: $\Theta = \frac{\theta(1-\theta)(1-\sigma_1)}{2\theta(1-\theta)(1-\sigma_1)+(1-\theta)^2}$, which is certainly less than $1/2$ irrespectively of the value of σ_1.

Therefore, we are *de facto* in the situation in which $\sigma(k) = 1$, for all k. In this case, as we proved when illustrating the chain of lies effect, egoists cannot be prevented from lying in the first period. Notice that the first-period principal has no instruments at his disposal for countering the incentive of a low benefit egoist to declare altruist, since, whatever the first period allocation, he is indifferent as to reporting his permanent type. So separation of altruists from egoists becomes impossible. This means that altruism is completely ineffective in the first period — or, more generally, in all periods except the last.

4.2 Altruism.

The definition of altruism we have adopted above in intertemporal decisions with mobility is such that in the first period altruists are concerned with the future fate of their first-period mates. Indeed, a different variety of altruism is conceivable: as far as future periods are concerned, in today's decisions an altruist cares about the future surplus not of today's but of tomorrow's fellows. This slight change has significant effects. Going back to the example presented above for describing the white lie effect, it is straightforward to check that the strategy we have examined in that context would become feasible. In fact, now for an altruist the shift from $\sigma(1) = 0$ to $\sigma(1) = 1$ generates no incentive to declare egoist: indeed his objective function is satisfied exactly at the same level irrespective of whether the group is stabilized or not, since in both cases group surplus takes on the efficiency level. Observe that in this example selective group stabilization is socially beneficial. On commenting this result one could notice that the initial definition of altruism is more "particularistic" in nature, so it is not surprising if under that initial definition altruists exert greater resistance to the principal's attempts at maximizing social surplus. Still, the "particularistic" definition seems to us more realistic as a description of concern for group welfare.

5 Conclusions

The model we presented helps clarify that, in hidden information contexts, the value of repeated interactions is not necessarily embodied in long-lasting fruitful relationships themselves but rather in the possibility of eliciting information about group composition in the first stage of interaction and using it for selectively stabilizing or terminating relationships. Suppose for instance that all the principals but one adopt the strategy of never stabilizing groups. Then the remaining principal may find it advantageous to stabilize groups with a relatively high share of altruists, and discontinue the others. This possibility is limited, but not necessarily excluded, by the agent's reluctance to reveal their degree of cooperativeness. In fact we find Nash equilibria with stabilization. The difference between the two statements above is not apparent in Ghosh and Ray's model, since their bad agents (i.e. the impatient) have "non cooperation" as the dominant strategy, so they self-select themselves away from good agents - who, on their turn, have nothing to gain from being considered bad. So all productive relationships (those with 2 patients players in their setting) persist in equilibrium. Instead, in our model, given that both types have good reasons for cheating, exploiting information requires that certain fruitful relationships be discontinued (in particular those with two altruists). Furthermore, the decision of one principal on group termination exerts externalities on the other principals, via the expected composition of fresh groups. Thus, whilst social surplus and group surplus coincide in Ghosh and Ray's model, in our model maximization of group surplus may entail the stabilization of some groups, despite social surplus is certainly less with stabilization than in the case of non repeated interactions.

Our model also sheds some light on the effects of altruism in social relationships. On the one hand, due to adverse intertemporal incentives connected with private information, the presence of even perfect altruists can be ineffective at improving social surplus when mobility decisions are left to individual discretion. On the other hand, the notion of altruism, that appears natural in the context of a single group, appears particularistic when it is applied to an intertemporal context with mobility, so altruism may results less beneficial to society than it is in a sequence of non repeated interactions.

References

Fudenberg D. and J. Tirole (1991), *Game Theory*, The MIT Press

Ghosh, P. and Ray, D. (1996) "Cooperation in Community Interaction without Information Flows", *Review of Economic Studies*, vol. 63, pp. 491-519.

Kranton R.E (1996), " The Formation of Cooperative Relationships", *Journal of Law Economics and Organization* 12 : 214-233.

Milgrom, P. and J. Roberts (1982) "Predation, reputation and entry deterrence", *Journal of Economic Theory*, vol. 27, pp. 280-312.

Rauch, J.E. and J. Watson (2003) "Starting Small in an Unfamiliar Environment", *International Journal of Industrial Organization*, vol. 21, pp. 1021-1042.

Sen A. (1985) "Goals, Commitment, and Identity", *Journal of Law, Economics and Organization*, vol. 1, pp. 341-355.

Watson J (2002), "Starting Small and Commitment", *Games and Economic Behavior* 38 : 176-199.

Appendix A

Proof of Fact 1.

When members' permanent types are common knowledge 3 cases are possible.

Case 1: both agents are of type a.

It is immediately seen that efficiency obtains at equilibrium. This means that $r(2ah) = r(ah.al) = 1, r(2al) = 0$. Moreover, the payments are as follows: $g(2ah, ah) = c/2$, $g(ah.al, ah) = c$, $g(ah.al, al) = 0$, $g(2al, al) = 0$.

Case 2: one agent is of type a and one of type e.

It is immediately seen that again efficiency obtains at equilibrium. This means that $r(ah.eh) = r(ah.el) = r(eh.al) = 1$, $r(al.el) = 0$. Moreover the payments are as follows: $g(ah.eh, ah) = g(ah.el, ah) = g(eh.el, eh) = c$, $g(eh.al, al) = g(2al, al) = 0$.

Case 3: both agent are of type e.

As, of course, at equilibrium it is $r(2eh) = 1$, $r(2el) = 0$, $g(2eh, eh) = c/2$, $g(eh.el, eh) = c$, and $g(eh.el, el) = 0$, in order for the incentive compatibility constraint for type eh to be satisfied, it must be

$$(1 - c/2)p + (1 - c)(1 - p)r(eh.el) \geq pr(eh.el) \tag{40}$$

Notice that from $p > 2\frac{1-c}{2-c(1+\theta)}$ (assumption a4'), it follows that: $1 - 2p - c + cp < 0$ and $-\frac{1}{2}(2 - c)\frac{p}{1-2p-c+cp} < 1$. Hence from (40) we get $r(eh.el) \leq -\frac{1}{2}(2 - c)\frac{p}{1-2p-c+cp} < 1$.

Next, let us consider the case in which individual permanent types are private information.

Equilibrium individual payments are easily characterized. When a member is high benefit and his mate is low benefit the former pays all the cost of provision. When both members are high benefit the cost is equally shared, except when one member is altruist and his mate is egoist in which case the former pays all the cost (this is for incentive purpose, as an altruist is indifferent as to who pays the cost, whereas an egoist is tempted to lie to avoid any payment). As regards the probabilities of provision, suppose that at equilibrium the binding incentive compatibility constraints are $IC(eh, el)$, $IC(eh, al)$, $IC(ah, eh)$ – we will verify this later. Therefore we have the following 3 equations:

$$\theta(p + (1 - p)(1 - c)r(eh.al)) + (1 - \theta)p\left(1 - \frac{1}{2}c\right) + \\ + (1 - \theta)(1 - p)(1 - c)r(eh.el) - \theta pr(ah.el) - (1 - \theta)pr(eh.el) = 0 \tag{41}$$

$$\theta\left(p+\left(1-p\right)\left(1-c\right)r(eh.al)\right)+\left(1-\theta\right)p\left(1-\frac{1}{2}c\right)+ \\ +\left(1-\theta\right)\left(1-p\right)\left(1-c\right)r(eh.el)-\theta pr(ah.al)-\left(1-\theta\right)pr(eh.al)=0 \quad (42)$$

$$\left(1-p\right)\left(1-c\right)\left(\left(1-\theta\right)r(ah.el)+\theta r(ah.al)\right)-\left(1-p\right)\left(1-c\right)\left(\left(1-\theta\right)r(eh.el)\ +\theta r(eh.al)\right)=0 \quad (43)$$

The solution to this system has one degree of freedom. So let us write $r(eh.al)$, $r(ah.el)$, $r(ah.al)$ as a function of $r(eh.el)$. It is:

$$r(eh.al)=r(ah.el)=\frac{2r(eh.el)(\theta+c-1-c\theta+2p-2p\theta-pc+pc\theta)-2p+pc-pc\theta}{2\theta\left(-2p+1-c+pc\right)} \quad (44)$$

and

$$r(ah.al)=\frac{2p-pc+3pc\theta-2pc\theta^2}{2\theta^2\left(-2p+1-c+pc\right)}+ \\ +\frac{2r(eh.el)(1+4p\theta+2c\theta+pc-2\theta-c-2p-2pc\theta+\theta^2-4\theta p-c\theta^2-2p\theta^2+pc\theta^2)}{2\theta^2\left(-2p+1-c+pc\right)} \quad (45)$$

Now let us verify that $IC(eh,ah)$

$$\theta\left(p+\left(1-p\right)\left(1-c\right)r(eh.al)\right)+\left(1-\theta\right)p\left(1-\frac{1}{2}c\right)+\left(1-\theta\right)\left(1-p\right)\left(1-c\right)r(eh.el)\geq \\ \theta p(1-c/2)+(1-\theta)p(1-c)+\theta(1-p)(1-c)r(ah.al)+\left(1-\theta\right)\left(1-p\right)\left(1-c\right)r(ah.el) \quad (46)$$

is not binding. By substituting (44) and (45) into (46), the left side becomes $-\frac{1}{2}p^2\frac{c\theta-c+2}{pc-2p+1-c}$ which is greater than the right side (which amounts to $-\frac{1}{2}p\frac{c-3pc-c^2+pc^2+2p+p\theta c}{pc-2p+1-c}$).

Next we show that $IC(al,el)$ is satisfied as an equality when $IC(eh,el)$ and $IC(eh,al)$ are binding. In fact, from

$$\theta\left(p+\left(1-p\right)\left(1-c\right)r(eh.al)\right)+\left(1-\theta\right)p\left(1-\frac{1}{2}c\right)+\left(1-\theta\right)\left(1-p\right)\left(1-c\right)r(eh.el)+ \\ -\theta pr(ah.el)-\left(1-\theta\right)pr(eh.el)=0 \quad (47)$$

and

$$\theta\left(p+\left(1-p\right)\left(1-c\right)r(eh.al)\right)+\left(1-\theta\right)p\left(1-\frac{1}{2}c\right)+\left(1-\theta\right)\left(1-p\right)\left(1-c\right)r(eh.el)+ \\ -\theta pr(ah.al)-\left(1-\theta\right)pr(eh.al)=0 \quad (48)$$

it follows that $\theta r(ah.al) + (1-\theta)\, r(eh.al) - (\theta r(ah.el) + (1-\theta)\, r(eh.el)) = 0$, which is $IC(al, el)$ written as an equality.

What remains to prove is our hypothesis that constraints (41)-(43) are binding. Let us add three slack variables u_1, u_2, and u_3, required to obtain equality. Then, compute again the system solutions which are now expressed as (cumbersome) functions of $r(eh.el)$ and of all the slacks. Finally, substitute them into the principal objective function and take first derivatives with respect to u_1, u_2, u_3. It results:

$$\frac{\partial S}{\partial u_1} = 2p^2\,(1-p)\,(1-c)\,\frac{1-\theta}{1+pc-c-2p} < 0 \tag{49}$$

$$\frac{\partial S}{\partial u_2} = 2p\,(1-p)\,(1-c)\,\theta\frac{(1-c)\,(1-p)}{1+pc-c-2p} < 0 \tag{50}$$

$$\frac{\partial S}{\partial u_3} = 2\,(1-p)\,(1-c)\,\frac{p}{1+pc-c-2p} < 0 \tag{51}$$

– recall that $1+pc-c-2p$ is negative by assumption a4'.

This proves that our hypothesis is true: at equilibrium $IC(eh, el)$, $IC(eh, al)$, $IC(ah, eh)$ are binding.

When the probabilities take the values defined in (44) and (45), it is $\frac{dS}{dr(eh.el)} = 0$, thus $r(eh.el)$ can really be set at discretion.

So let us set it to its equilibrium value in a group where it is common knowledge that there are no altruists, that is $r(eh.el) = p\frac{1-c/2}{p-(1-p)(1-c)}$. Then it is immediately verified that $r(eh.al) = r(ah.el) = \frac{p}{2p-1+c-pc} < 1$ for $p < 1$, $r(ah.al) = \frac{1}{2}\frac{-c+c\theta+2\theta}{\theta} r(ah.el) < 1$ since it is both $r(ah.el) < 1$ and $\frac{-c+c\theta+2\theta}{\theta} = (2+c) - c/\theta < 2$.

Therefore, we conclude that expected group surplus is lower when individual permanent types are private information than when they are common knowledge.

Appendix B.

This appendix proves that when $N = 1$, $\sigma(0) = \sigma(1) = \sigma(2) = 1$ cannot occur at equilibrium.

When $\sigma(0) = \sigma(1) = \sigma(2) = 1$ the second-period expected surplus of an egoist if he reports a (and then he optimally declares l) is p^2, while if he truthfully reports e it is:

$$p\,\{\theta\,(p + (1-p)(1-c)) + (1-\theta)\,[p\,(1-c/2) + (1-p)(1-c)\nu]\} \tag{52}$$

where $\nu = r(eh.el)$ is the probability of providing the good when there are two egoists with opposite evaluations in the s_0-mechanism. Let $\overline{\nu}$ denote the highest value of ν such that lying is preferable for an egoist, so $\overline{\nu}$ solves $p\theta\,(p + (1-p)(1-c)) + p\,(1-\theta)\,[p\,(1-c/2) + (1-p)(1-c)\overline{\nu}] -$

$p^2 = 0$. Hence this threshold amounts to

$$\overline{\nu} = \frac{1}{2}\frac{2\theta(c-1) - 3\theta pc + pc + 2\theta p}{(1-c)(1-p)(1-\theta)} \tag{53}$$

On the other hand, we know, from Appendix A, that in the equilibrium of the s_0-mechanism it is $\nu = \frac{1}{2}p\frac{-2+c}{1-c-2p+pc}$[12].

So, if $\overline{\nu} - \nu > 0$ we know that the intertemporal incentive constraint of a type-e is violated. Let us check this point. We have that

$$\overline{\nu} - \nu = -\frac{1}{2}\frac{p^2\,(2-c)\,(1+\theta-2\theta c) - 2p\,(1-c)\,(1+2\theta-2\theta c) + 2\theta\,(1-c)^2}{(1-\theta)\,(1-p)\,(1-c)\,(1-c-2p+pc)} \tag{54}$$

The denominator is obviously negative. The numerator is a U-shaped parabola in p. Since the first and the last coefficient are positive, while the middle is negative, if there are two intersections with the $p-$axis, these are positive. The largest is $\overline{p} = (1-c)\,\frac{(1+2\theta-2\theta c)+\sqrt{(1-2\theta c+2\theta^2 c)}}{(2-c)(1+\theta-2\theta c)}$. If $p > \overline{p}$ then $\overline{\nu} - \nu > 0$. Now we show that $\overline{p}$ is smaller than the threshold stated by assumption a4', i.e. $\overline{p} \leqslant \frac{2-2c}{2-c(1+\theta)}$.

$$\begin{aligned}\overline{p} - \frac{2-2c}{2-c(1+\theta)} = (1-c)\left[\frac{(1+2\theta-2\theta c)+\sqrt{(1-2\theta c+2\theta^2 c)}}{[2-c(1+\theta)]+2\theta\,(1-c)^2} - \frac{2}{2-c(1+\theta)}\right] < \\ (1-c)\left[\frac{(1+2\theta-2\theta c)+\left(1-\theta c+\theta^2 c\right)}{(2-c)\,(1+\theta-2\theta c)} - \frac{2}{2-c(1+\theta)}\right] = \\ -(1-c)\theta c^2\frac{(1-\theta)^2}{(2-c)\,(1+\theta-2\theta c)\,[2-c(1+\theta]} < 0\end{aligned} \tag{55}$$

where the first inequality derives from the fact that $\sqrt{\left(1-2\theta c+2\theta^2 c\right)} < 1-\theta c+\theta^2 c$. [13]

So we can conclude that, in the relevant range of p, it is $\overline{\nu} - \nu > 0$, i.e. the incentive constraint of egoists is always violated.

---

[12] In the equilibrium of the s_0-mechanism it is $p\,(1-c/2) + (1-p)(1-c)\nu = \nu p$, whence: $\nu = \frac{1}{2}p\frac{-2+c}{1-c-2p+pc}$. Notice that the denominator is negative, since $p > 1-c$, as implied by the condition for inefficiency.

[13] Notice that $1 > \sqrt{\left(1-2\theta c+2\theta^2 c\right)} > \left(1-2\theta c+2\theta^2 c\right) > 0$
In fact $1 > 1-2\theta c+2\theta^2 c$ since $\theta < 1$ and $1-2\theta c(1-\theta) > 0$, since $\theta(1-\theta) < 1/2$.
We know that for $0 < q < 1$, $\sqrt{q} < (1+q)\,/2$ (this is implied by $(1-\sqrt{q})^2 > 0$). So
$\sqrt{\left(1-2\theta c+2\theta^2 c\right)} < \frac{2-2\theta c+2\theta^2 c}{2} = 1-\theta c+\theta^2 c$.