A Large Population Approach to Implementing Efficiency with Minimum Inequality

Sarvesh Bandhu* and Ratul Lahkar†

September 15, 2022

Abstract

We consider the implementation of efficiency with minimum inequality in a large population model of negative externalities. Formally, the model is one of tragedy of the commons with the aggregate strategy at the efficient state being lower than at the Nash equilibrium. A planner can restore efficiency by imposing an externality equivalent tax and then redistribute the tax revenue as transfers to lower inequality. We characterize the transfer vector that minimizes inequality at the efficient state subject to incentive compatibility and budget balance. We then construct a mechanism that implements efficiency with minimum inequality in dominant strategies. We also show that minimizing inequality at the efficient state maximizes the minimum payoff at efficiency. But it is not equivalent to implementing the Rawlsian social choice function.

Keywords: Negative Externalities; Efficiency; Equality; VCG mechanism; Implementation.

JEL classification: C72; D62; D63; D82.

*Economics and Social Sciences, Indian Institute of Management, Bangalore, Bannerghatta Main Road, Bangalore, Karnataka 560076, India, email: sarvesh.bandhu@iimb.ac.in.

†(corresponding author) Department of Economics, Ashoka University, Rajiv Gandhi Education City, Sonipat, Haryana 131029, India, email: ratul.lahkar@ashoka.edu.in.

1 Introduction

The classical literature on welfare economics and mechanism design has considered the question of achieving social efficiency in great detail. Efficiency is generally interpreted as implementing the utilitarian social choice function that seeks to maximize aggregate payoff in society. Perhaps the most well–known mechanisms pursuing this objective is the class of Vickrey–Clarke–Groves (VCG) mechanisms (Vickrey [24], Clarke [3], Groves [8]), which take a dominant strategy implementation approach to this problem. As far as we know, however, this literature has not addressed one important question. That question is one of minimizing inequality at the socially efficient state subject, of course, to the usual incentive compatibility and budget balance constraints. One reason may be that in the usual setting of finite player mechanisms, dominant strategy implementation of efficiency with incentive compatibility is incompatible with budget balance (Green and Laffont [7]).[1] Adding the objective of minimum inequality would only make the problem even more intractable. But equality is an important objective for any society. This is not just on grounds of fairness but also because higher inequality may prove detrimental to the utilitarian objective of economic growth by, for example, facilitating elite capture of institutions (Sokoloff and Engerman [21]). Therefore, reconciling efficiency with equality is a significant question.

We consider precisely this question in this paper. But since a finite player setting would prove intractable, we adopt a large population approach where there are a continuum of agents, each agent being of measure zero. Unlike in finite player mechanisms, we can accommodate goals of efficiency with incentive compability and budget balance in large population mechanisms (Lahkar and Mukherjee [11]). Therefore, it may also be possible to combine minimum inequality with efficiency in such an environment, which is what this paper seeks to establish. Existing models of implementation in large population games focus exclusively on efficiency. Thus, Sandholm [17, 18, 19], who pioneered this literature, and Lahkar and Mukherjee [10, 12] consider the evolutionary implementation of efficiency in large population games.[2] Our broader objective of efficiency with minimum inequality, however, makes the evolutionary implementation approach inapplicable. Instead, we need to rely upon a more conventional dominant strategy implementation approach. Our approach has similarities to the large population VCG mechanism in Lahkar and Mukherjee [11] which implements efficiency in strictly dominant strategies in a public goods model. But again, that model does not consider equality issues. Thus, to the best of our knowledge, whether with a finite number or with a large population of players, this paper is the first to consider implementing a social choice function that combines efficiency with minimum inequality.

We consider a model with strategic interlinkages and negative externalities. Total output is a

[1] Thus, VCG mechanism implements efficiency by leaving a budget surplus with the planner. Budget balance would require that there should be neither surplus nor deficit with the planner. An alternative to the VCG mechanism is the AGV mechanism (Arrow [1], d'Aspremont and Gérard–Varet [5]) that generates truthful revelation as a Bayesian Nash equilibrium. But this mechanism requires the stronger assumption that the type distribution is common knowledge and also does not satisfy ex-post individual rationality.

[2] In evolutionary implementation, the planner imposes an externality price to generate a potential game (Sandholm [16]). The original efficient state then becomes the Nash equilibrium of the potential game to which, evolutionary dynamics converge. We note that there is no reliance on truthful revelation in such a model.

function of aggregate effort by agents, which is the source of these interlinkages. Different types of agents have different effort cost functions, which is private information, and they receive a share of total output according to individual effort. Thus, payoffs are equivalent to that of a tragedy of the commons model in which higher aggregate strategy reduces payoff (Lahkar and Mukherjee [12]). Hence, there are negative externalities. The tragedy of the commons is, of course, a canonical model of negative externalities in economics. It becomes relevant if the production process has the two characteristics of a common resource, non–excludability and rivalry. Consider, for example, the financial sector that produces exotic new financial products. No financial institution can be excluded from developing such products. But as more such institutions develop and start holding each other's products thereby creating interlinkages, the possibility of one player's actions having catastrophic consequences on others increases as exemplified by the 2008 financial crisis.[3] In our model, as in any tragedy of the commons, the consequences of the negative externality arises in the form of the aggregate strategy level at the Nash equilibrium being too high relative to the aggregate strategy level at the socially efficient level, or the level that maximizes aggregate payoff.

The discrepancy between the Nash equilibrium and social efficiency creates scope for benign intervention by a planner to restore efficiency. The planner can do so by taxing the externality causing activity. This is also the approach of the evolutionary implementation models of Sandholm [17, 18, 19] and Lahkar and Mukherjee [10, 12].[4] Similarly, Lahkar and Mukherjee [11] also apply externality pricing in their large population public goods model to implement efficiency in strictly dominant strategies. We, however, go further than these papers and propose a transfer scheme that redistributes the tax revenue in a way that reduces inequality without sacrificing efficiency. Thus, a novel feature of this paper is to design a tax and transfer scheme that simultaneously resolves the problem of negative externalities and improves equality. From a policy perspective, this is an important finding. It illustrates that improving efficiency and enhancing equality are not contradictory goals. Instead, the resources the policy maker requires to promote equality can arise from the taxation of those very activities that harm efficiency.[5] Analytically, our approach can also be extended to a model of positive externalities like the public goods game in Lahkar and Mukherjee [11], although we do not present such an extension here. The problem there would be to characterize a vector of inequality minimizing taxes that provides the revenue to subsidize a socially beneficial activity and restore efficiency. Considering negative externalities, though, allows us to focus on a different type of policy challenge; taxing a harmful activity and then seeking the

[3]See, for example, Elliot et al. [6] for an analysis of such financial interlinkages from a network theory perspective. Also see Chakrabarti and Lahkar [2] for a discussion of, besides finance, other industries like railways and information technology that may be modeled as large population tragedies of the commons due to the nature of interlinkages between them. In all such examples, the productive opportunities that these industries are seeking to exploit are being interpreted as the common resource.

[4]Sandholm [17, 18] are deterministic evolutionary implementation models whereas Sandholm [19] is one of stochastic evolution. These models consider large population games with a finite strategy set. Lahkar and Mukherjee [10, 12] extend the deterministic method to games with a continuous strategy set and apply it to problems like public goods, public bads and tragedy of the commons. The model in this paper is also a continuous strategy one.

[5]Thus, going back to our illustrative example, it may be possible to curb the speculative excesses of the financial sector through a financial transactions tax and use the proceeds for redistributive purposes.

inequality minimizing vector of redistributive transfers

Formally, we measure inequality as the variance of agents' payoffs at the efficient state following taxes and transfers. The planner seeks to minimize this variance subject to incentive compatibility and budget balance. We design a variant of the classical VCG mechanism suitably adapted to our large population context. Based on reported types, the planner assigns strategies, taxes and transfers to agents. We characterize the transfers that make truthful revelation weakly dominant while minimizing the variance of payoffs at the efficient state. Due to considerations of incentive compatibility, this variance is not zero. Hence, the equality achieved is not perfect. Nevertheless, agents disadvantaged with a higher cost of effort still receive a higher transfer due to which the inequality that remains is less than that achievable through, for example, an equal redistribution for all agents. Intuitively, equal redistribution, which suffices for efficiency, can be implemented with truthful revelation being strictly dominant. This leaves the planner enough scope to adjust incentive compatibility conditions so as to design the optimal transfer scheme that makes truthful revelation weakly dominant and thereby improve equality without compromising efficiency.

An important technical caveat to our results is that they hold for large population models or models where all agents are of measure zero. This is an important assumption because our analysis relies on the fact that changes in individual strategy does not affect aggregate variables. This adds considerably to the tractability of our problem. For example, the budget balance condition, which is crucial for us, is satisfied at efficiency in our large population context. But, as noted earlier, it is difficult to achieve in conventional finite player mechanisms. Of course, in real world situations, no agent is ever of measure zero. But in most economic environments where public policy questions like redistribution assume importance, we would expect the number of people involved to be fairly large. Further, it is reasonable to assume that in such situations, agents would behave as if their individual actions cannot influence aggregate variables. In that case, as in models of competitive markets, we would expect our conclusions to be valid, at least approximately.

Independent of efficiency, the classical implementation literature has considered equality from the point of view of implementing the Rawlsian social choice function (Rawls [15]), which seeks to maximize the minimum welfare in society. It is known, for example, that this social choice function is not implementable. While this is not our main focus, our model will also provide some insights into combining utilitarian and Rawlsian objectives. In particular, we show that subject to the feasibility constraints, our transfer scheme not only minimizes inequality at efficiency but also maximizes the minimum payoff at the efficient state. Thus, in this sense, our transfer scheme is a Rawlsian one but restricted to the efficient state. It provides a partial reconciliation of the utilitarian objective of efficiency with the Rawlsian objective of maximizing the minimum payoff. The reconciliation, however, is not complete because by using a counterexample, we show that this is not equivalent to implementing the Rawlsian social choice function. If we are willing to sacrifice efficiency, then we can identify another state where the minimum payoff is higher but aggregate payoff is lower. This is the Rawlsian outcome subject to incentive compatibility and budget balance. Again, as far as we know, the existing literature on implementation theory has not addressed such differences between implementing the Rawlsian social choice function and implementing the efficient outcome with minimum inequality.

There are also papers that address the issue of fairness at efficiency in other ways. A widely recognized axiom underlying fairness is envy freeness. No agent should desire the allocation of another agent. There are both positive and negative results on whether envy freeness is consistent with efficiency. For example, models by Tadenuma and Thomson [23] and Pápai [14] highlight conflict between envy freeness and other goals like efficiency, incentive compatibility and budget balance. On the other hand, Ohseto [13] and Sprumount [22] present models where efficiency and envy freeness are compatible. The details of these models are very different from ours. They are finite player object allocation models while ours is a large population model of negative externalities. Hence, direct comparison of results are difficult. Nevertheless, it is easy to see that our model does satisfy envy freeness at the efficient state. Transfers are designed in such a way that no agent prefers the allocation of any other type of agents.

The rest of the paper is as follows. Section 2 presents our model of the tragedy of the commons and characterizes its Nash equilibrium and efficient state. In Section 3, we identify the transfer vector that minimizes inequality at the efficient state subject to budget balance and incentive compatibility. Section 4 describes the mechanism that implements efficiency with minimum inequality in dominant strategies. Section 5 presents the counterexample about the Rawlsian social choice function. Section 6 concludes.

2 The Model

We consider a society consisting of a continuum of agents, each of measure zero. The society is divided into a finite set of populations, also called types, $\mathcal{P} = \{1, 2, \cdots, n\}$. The mass of type $p \in \mathcal{P}$ is $m_p \in (0, 1)$ with $\sum_{p\in\mathcal{P}} m_p = 1$. Thus, the total mass of the society is 1. We refer to the distribution $m = (m_1, m_2, \cdots, m_n)$ as the type distribution in the society. Every agent in the society has a common strategy set $\mathcal{S} = (0, \infty)$.[6] Throughout, we will interpret $x \in \mathcal{S}$ synonymously as the effort exerted by an agent. Let $\mathcal{M}^+_\lambda(\mathcal{S})$ be the space of finite signed measures that impose a mass $\lambda > 0$ on $\mathcal{S}$. We then use the measure $\mu_p \in \mathcal{M}^+_{m_p}(\mathcal{S})$ to denote the state of population p. The population state describes the strategy distribution in that population. Thus, $\mu_p(A) \in [0, m_p]$ is the mass of agents in population p who are playing strategies in $A \subseteq \mathcal{S}$. If every agent in population p plays the same strategy x, then we obtain a monomorphic population state which we denote it as $m_p\delta_x$. We interpret the vector of population states $\mu = (\mu_1, \mu_2, \cdots, \mu_n) \in \Delta = \prod_{p\in\mathcal{P}} \mathcal{M}^+_{m_p}(\mathcal{S})$ as the state of the entire society or the social state. The aggregate strategy level in the society at the social state μ is then

$$A(\mu) = \sum_{p\in\mathcal{P}} \int_{\mathcal{S}} x\mu_p(dx). \tag{1}$$

In our subsequent analysis, we will frequently refer to $A(\mu) \in (0, \infty)$ as α.

[6] For certain technical reasons explained later, we exclude the 0 strategy.

We consider an economy with interlinkages between agents. To capture such interlinkages, we assume that total output in the society depends upon the aggregate strategy level. Formally, we consider a smooth, strictly increasing and strictly concave production function $\pi : (0, \infty) \to \mathbf{R}_+$ such that $\pi(A(\mu))$ is the total output in the society when the aggregate strategy is $A(\mu)$. We assume that π satisfies the Inada conditions and that $\frac{\pi(\alpha)}{\alpha}$ is strictly declining for all $\alpha \in (0, \infty)$. An agent exerting effort x then receives a share $\frac{x}{A(\mu)}$ of the total output and incurs an effort cost $c_p(x)$. These cost functions are the source of type specific distinctions in our model. They differ according to the type of agents but it is the same for all agents of a particular type. We assume that every such type specific cost function $c_p : \mathcal{S} \to \mathbf{R}_+$ is smooth, strictly increasing and strictly convex and satisfies $c_p(0) = c_p'(0) = 0$ if we extend the function to 0, where $c_p'(x)$ is type p's marginal cost. Thus, there are no fixed costs and the marginal cost also tends to zero as $x \to 0$. In addition, we make the following assumption about the cost functions.

Assumption 2.1 *For every $p, q \in \mathcal{P}$, $c_q(x) - c_p(x)$ is strictly increasing in x if $q > p$.*

This assumption can be equivalently written as $c_q'(x) > c_p'(x)$ for all $x \in \mathcal{S}$ if $q > p$, i.e. marginal cost at any level of effort is higher for higher types. It has an important implication. Recall that fixed cost is zero for all types. Hence, the area beneath the marginal cost is total cost. Assumption 2.1, therefore, generates the following observation.

Observation 2.2 *For every $x \in \mathcal{S}$, $c_p(x)$ is strictly increasing in $p \in \mathcal{P}$. Thus, for every $x \in \mathcal{S}$, $c_1(x) < c_2(x) < \cdots < c_n(x)$.*

This observation gives us an important labeling convention in our model. Higher cost agents are classified as higher types. Hence, we may interpret agents labeled as being of a higher type as facing a greater disadvantage in exerting effort.

A population game is a weakly continuous mapping $F : \mathcal{S} \times \mathcal{P} \times \Delta \to \mathbf{R}$ such that $F_{x,p}(\mu)$ is the payoff of an agent from population p who plays strategy x at the social state μ. Given the production and cost functions, this payoff in our model takes the form

$$\begin{aligned} F_{x,p}(\mu) &= \frac{x}{A(\mu)}\pi(A(\mu)) - c_p(x) \\ &= xAP(A(\mu)) - c_p(x), \end{aligned} \tag{2}$$

where $AP(A(\mu)) = \frac{\pi(A(\mu))}{A(\mu)}$ is the average product of the production function when aggregate effort is $A(\mu)$. Thus, our assumption that $\frac{\pi(\alpha)}{\alpha}$ is strictly declining is equivalent to the average product function being strictly declining. Formally, the payoff (2) is equivalent to a large population tragedy of the commons model with the aggregate output $\pi(A(\mu))$ being shared among agents in proportion to their individual effort x (Lahkar and Mukherjee [12]).[7] We should note, however, that the validity of our model doesn't depend upon the existence of a literal physical common resource. All we need

[7]See footnote 6. The reason for excluding the 0 strategy is to ensure that (2) is defined at all social states. Otherwise, if all agents play 0, the average product would be undefined.

is that total output should depend upon aggregate strategy which is then shared among agents. Non–excludability arises because no agent can be stopped from contributing to aggregate strategy. Rivalry arises from diminishing average product.

2.1 Nash Equilibrium and Efficient State

The population game F defined by (2) is an aggregative game as the payoff of an agent depends entirely upon his individual strategy and the aggregate strategy level $A(\mu)$ (Corchón [4]). We now use this aggregative structure of F to characterize its Nash equilibrium and efficient state.[8]

Let us denote the aggregate strategy level $A(\mu)$ as α and write (2) as $xAP(\alpha) - c_p(x)$. The strict convexity of $c_p(x)$ implies that for every given α, this function has a unique maximizer in $\mathcal{S}$. This maximizer, which we denote as $b_p(\alpha)$, is the unique best response of a type p agent to every social state μ such that $A(\mu) = \alpha$. The following proposition then characterizes the unique Nash equilibrium of our model. Further details of the proof are in Appendix A.1.[9]

Proposition 2.3 *Consider the population game F defined by (2). Denote by α^N the unique solution to*

$$\sum_{p\in\mathcal{P}} m_p b_p(\alpha) = \alpha. \tag{3}$$

Then, F has a unique Nash equilibrium

$$\mu^N = \left(m_1\delta_{\alpha_1^N}, m_2\delta_{\alpha_2^N}, \cdots, m_n\delta_{\alpha_n^N}\right) \tag{4}$$

where $\alpha_p^N = b_p(\alpha^N)$ and $b_p(\alpha)$ is the unique best response function in F as characterized in (5). Thus, every agent of type $p \in \mathcal{P}$ plays strategy $\alpha_p^N = b_p(\alpha^N)$ at this Nash equilibrium. The aggregate strategy at μ^N is, therefore, $\alpha^N = \sum_{p\in\mathcal{P}} m_p\alpha_p^N$. The Nash equilibrium is characterized by

$$AP(\alpha^N) = c_p'(b_p(\alpha^N)). \tag{5}$$

Intuitively, (3) implies that a Nash equilibrium of an aggregative game is a social state such that when all agents play their best response to that state, the aggregate strategy level remains unchanged. The key to Proposition 2.3 is that $b_p(\alpha)$ is strictly declining due to our assumptions about $AP(\alpha)$ and the cost functions. Hence, (3) has a unique solution in our model, which characterizes the unique Nash equilibrium.

Condition (5) implies that this Nash equilibrium involves equating average product to marginal cost. That cannot be efficient. Instead, to characterize the efficient state, we consider the aggregate

[8] These results have also been established in Lahkar and Mukherjee [12]. Nevertheless, we present them here as well briefly in order to keep the present paper self–contained. Moreover, due to their focus on evolutionary implementation, Lahkar and Mukherjee [12] apply the method of potential games to derive these results (see footnote 2). Potential games do not play any role in the implementation approach of this paper. Instead, we use the aggregative structure of our model and apply more direct methods relying on best responses to derive our results. Also see Lahkar [9] for an application of such methods to aggregative games.

[9] The superscript N in Proposition 2.3 indicates "Nash".

payoff. The aggregate payoff in a population game F at a social state μ, denoted $\bar{F}(\mu)$, is the total payoff earned by all agents at that state. Hence, given the payoff function (2), the aggregate payoff in our model is

$$\begin{aligned}\bar{F}(\mu) &= \sum_{p\in\mathcal{P}} \int_{\mathcal{S}} F_{x,p}(\mu)\mu_p(dx) \\ &= \sum_{p\in\mathcal{P}} \int_{\mathcal{S}} \left(\frac{x}{A(\mu)}\pi(A(\mu)) - c_p(x)\right)\mu_p(dx) \\ &= \frac{\pi(A(\mu))}{A(\mu)} \sum_{p\in\mathcal{P}} \int_{\mathcal{S}} x\mu_p(dx) - \sum_{p\in\mathcal{P}} \int_{\mathcal{S}} c_p(x)\mu_p(dx) \\ &= \pi(A(\mu)) - \sum_{p\in\mathcal{P}} \int_{\mathcal{S}} c_p(x)\mu_p(dx), \end{aligned} \tag{6}$$

where the last equality follows from the definition of the aggregate strategy in (1). Thus, the aggregate payoff at a state μ is the total output generated by the society at that state minus the aggregate cost $\sum_{p\in\mathcal{P}} \int_{\mathcal{S}} c_p(x)\mu_p(dx)$ incurred by agents at that state. An *efficient state* of F is then a state μ^* *that maximizes* the aggregate payoff (6) in Δ.

The strategic interlinkages in our model imply there are externalities. Therefore, characterizing an efficient state would require us to account for such externalities. Let $e_{x,p}(\mu)$ denote the total externality imposed by an agent of type p who plays strategy x at the state μ on the entire society. Corollary 5.7 in Lahkar and Mukherjee [12] calculates this total externality in a tragedy of the commons model such as the present one to be

$$e_{x,p}(\mu) = x\left(MP(A(\mu)) - AP(A(\mu))\right), \tag{7}$$

where $MP(\alpha) = \pi'(\alpha)$ is the marginal product of π at the aggregate strategy α. Our assumption that $AP(\alpha)$ is strictly declining implies $MP(\alpha) < AP(\alpha)$ at all $\alpha \in (0, \infty)$ so that $e_{x,p}(\mu) < 0$. Hence, externalities are negative in our model, which is another standard characteristic of tragedy of the commons problems.[10]

It is known from Sandholm [16] that an efficient state of a population game F is also a Nash equilibrium of another game $\hat{F}$ we obtain by adding externalities in F to the original payoffs. We interpret the addition of this externality as the imposition of a tax which compels agents to internalize the externality they create. The payoff of a type p agent who plays strategy x in $\hat{F}$ is

$$\begin{aligned}\hat{F}_{x,p}(\mu) &= F_{x,p}(\mu) + e_{x,p}(\mu) \\ &= xAP(A(\mu)) - c_p(x) + x(MP(A(\mu)) - AP(A(\mu))) \\ &= xMP(A(\mu)) - c_p(x). \end{aligned} \tag{8}$$

[10]See Appendix A.1.1 in Lahkar and Mukherjee [10] for the technical details of calculating externalities in large population games with a continuous strategy set. Also see Proposition 4.1 in Lahkar and Mukherjee [12] for a general derivation of externalities in aggregative games.

Like (2), (8) is also an aggregative game with the only difference being that the average product gets replaced by the marginal product. Hence, we can apply the same method as in Proposition 2.3 to obtain the Nash equilibrium of $\hat{F}$ or, equivalently, the efficient state of F. Thus, let $\hat{b}_p(\alpha)$ be the unique best response of a type p agent in $\hat{F}$ defined by (8) at a social state μ such that $A(\mu) = \alpha$. We then obtain the following result. Further details are in Appendix A.1.

Proposition 2.4 *Consider the population game $\hat{F}$ defined by (8) and the best response $\hat{b}_p(\alpha)$ characterized in (11). This game has a unique Nash equilibrium*

$$\mu^* = \left(m_1\delta_{\alpha_1^*}, m_2\delta_{\alpha_2^*}, \cdots, m_n\delta_{\alpha_n^*}\right), \tag{9}$$

where $\alpha_p^* = \hat{b}_p(\alpha^*)$ is the strategy of every agent of type p at μ^* and α^* is the unique solution to

$$\sum_p m_p\hat{b}_p(\alpha) = \alpha. \tag{10}$$

Hence, μ^* is also the efficient state of the original game F defined by (2). The aggregate strategy at μ^* is $\alpha^* = \sum_{p\in\mathcal{P}} m_p\alpha_p^*$. Further, for each p, $\alpha_p^* < \alpha_p^N$, the Nash equilibrium strategy level characterized in Proposition 2.3. Hence, $\alpha^* < \alpha^N$. Moreover, μ^* is characterized by

$$MP(\alpha^*) = c_p'\left(\hat{b}_p(\alpha^*)\right). \tag{11}$$

As is any model of negative externalities, the efficient state involves a lower strategy level than at the Nash equilibrium. This is true for all types of agents and, therefore, at the aggregate level as well. At the efficient state, as implied by (11), every agent equates the marginal product of π to his type specific marginal cost. This is, of course, the hallmark of efficiency. The following corollary provides a ranking of payoffs at the efficient state. Quite intuitively, agents with lower levels of cost are better off. They also exert higher effort in at the efficient state. The proof of the corollary is in Appendix A.1.

Corollary 2.5 *Consider the efficient state μ^* characterized in Proposition 2.4. Using (8) and Proposition 2.4, let us denote*

$$\hat{F}_{\alpha_p^*,p}(\mu^*) = \alpha_p^* MP(\alpha^*) - c_p(\alpha_p^*), \tag{12}$$

as the payoff of a type p agent at the efficient state of the tragedy of the commons F defined by (2). If $p < q$, then $\hat{F}_{\alpha_p^*,p}(\alpha^*) > \hat{F}_{\alpha_q^*,q}(\alpha^*)$. Moreover, $\alpha_p^* > \alpha_q^*$ i.e. $\alpha_1^* > \alpha_2^* > \cdots > \alpha_n^*$.

3 Efficiency and Equality

Henceforth, our focus is on the efficient state μ^* characterized in Proposition 2.4. We now envisage a planner who wishes to implement the efficient state but with minimum possible inequality. In

Section 4, we will discuss the formal methodology of implementing this objective. Here, we discuss certain preliminary issues that will be relevant for implementation. Suppose the efficient state μ^* has been achieved with every agent playing α_p^* as characterized in Proposition 2.4 and paying a tax $\alpha_p^* \left(AP(\alpha^*) - MP(\alpha^*)\right)$ equal (in absolute value) to the negative externality (7) they create at the efficient state (Proposition 2.4).[11] Hence, the total tax revenue the planner obtains at the efficient state is

$$T(\mu^*) = \sum_p m_p \alpha_p^* \left(AP(\alpha^*) - MP(\alpha^*)\right) = \alpha^* \left(AP(\alpha^*) - MP(\alpha^*)\right). \tag{13}$$

We now allow the planner to redistribute the entire tax revenue received among the agents as transfers. Notice from (8) that once the tax is paid, the payoff of every type p agent at the efficient state is only $\alpha_p^* AP\left(\alpha^*\right) - c_p\left(\alpha_p^*\right) + \alpha_p^*\left(MP(\alpha^*) - AP(\alpha^*)\right) = \alpha_p^* MP\left(\alpha^*\right) - c_p\left(\alpha_p^*\right)$, which is $\hat{F}_{\alpha_p^*,p}(\mu^*)$ as defined in (12). Due to (13), redistribution ensures that the entire aggregate payoff at the efficient state μ^*,

$$\begin{aligned}
&\sum_p m_p \left(\alpha_p^* MP\left(\alpha^*\right) - c_p\left(\alpha_p^*\right)\right) + T(\mu^*) \\
=&\alpha^* MP(\alpha^*) - \sum_p m_p c_p\left(\alpha_p^*\right) + T(\mu^*) \\
=&\alpha^* AP(\alpha^*) - \sum_p m_p c_p\left(\alpha_p^*\right) \\
=&\pi(\alpha^*) - \sum_p m_p c_p\left(\alpha_p^*\right) \\
=&\pi(A(\mu^*)) - \sum_{p\in\mathcal{P}} \int_{\mathcal{S}} c_p(x)\mu_p^*(x),
\end{aligned} \tag{14}$$

accrues to the agents. Note from (6) that (14) is just $\bar{F}(\mu^*)$.

Throughout, we assume that during the redistribution exercise, the planner provides the same transfer to every agent of a particular type, although the transfer may vary across types. Let $t = (t_1, t_2, \cdots, t_p)$ be a vector of such type specific transfers. We also assume that any such transfer vector satisfies the *budget balance* condition

$$\sum_p m_p t_p = T(\mu^*), \tag{15}$$

where $T(\mu^*)$ is as defined in (13). Then, at μ^*, if a type p agent plays his type specific efficient strategy α_p^*, we can use (12) to write his post redistribution payoff as

$$\hat{F}_{\alpha_p^*,p}(\mu^*) + t_p = \alpha_p^* MP\left(\alpha^*\right) - c_p\left(\alpha_p^*\right) + t_p \tag{16}$$

[11] A slight clarification about notation. When we interpret the tax as a payment from the agent to the planner, we write it as the positive amount $\alpha_p^*(AP(\alpha^*) - MP(\alpha^*))$. When we interpret it as a payment from the planner to the agent as in (16) below, we write it as the negative amount $\alpha_p^*(MP(\alpha^*) - AP(\alpha^*))$.

Equivalently, using (8), we can think of the post redistribution payoff as the original payoff $\alpha_p^* AP(\alpha^*) - c_p(\alpha_p^*)$ plus the *payment* $\left[\alpha_p^*(MP(\alpha^*) - AP(\alpha^*)) + t_p\right]$ made by the planner to a type p agent. The planner would like to also enhance equality during redistribution through these transfers. The budget balance condition implies that whatever resources the planner needs to promote equality comes from the taxation of the negative externality. Hence, taxation in our model will not only curb the negative externality but will also reduce inequality.

We measure inequality at the efficient state as the variance of these post redistribution payoffs (16). Budget balance ensures that the aggregate payoff once all agents receive (16) is $\bar{F}(\mu^*)$ as calculated in (14). Further, due to the measure zero characteristic of each agent, this aggregate payoff is also the average payoff in the society following redistribution. Therefore, the variance of the redistributed payoffs (16) at the efficient state of the tragedy of the commons (2) is

$$V(\mu^*, t) = \sum_{p \in \mathcal{P}} m_p \left[\alpha_p^* MP(\alpha^*) - c_p(\alpha_p^*) + t_p - \bar{F}(\mu^*)\right]^2. \tag{17}$$

The planner's objective is to choose a transfer scheme t that minimizes (17), which we shall denote as $t^* = (t_1^*, t_2^*, \cdots, t_n^*)$.

In the rest of this section, we focus on characterizing the vector t^*. For this purpose, we will use our earlier assumption made at the beginning of this section that the society is already at the efficient state μ^*. We also assume that the that the planner knows the type distribution m. It is legitimate to make such assumptions because the objective of this section is not to implement μ^* or t^*. Instead, our aim right now is the entirely technical one of characterizing t^*. In the next section, where we discuss the planner's actual implementation of both μ^* and t^*, we will drop these assumptions.

The transfer vector we seek will have to respect incentives for truthful revelation. To formalize those incentives, suppose the planner asks every agent to report his type. If an agent reports type to be q, then the planner assigns that agent the strategy α_q^* and the transfer t_q. The planner is able to assign α_q^* because we have assumed that he knows m and, therefore, can calculate the type specific efficient strategies $(\alpha_1^*, \cdots, \alpha_n^*)$ as in Proposition 2.4. We can then use (16) to write the payoff of a type p agent who claims to be of type q under a transfer scheme $t = \{t_1, t_2, \cdots, t_n\}$ as

$$\phi_p(q, \mu^*, t) = \alpha_q^* MP(\alpha^*) - c_p(\alpha_q^*) + t_q, \tag{18}$$

Hence, the payoff from truthful revelation, $\phi_p(p, \mu^*, t)$, is the same as (16).[12] Notice that despite a possible false report by this particular agent, the aggregate strategy in (18) is still α^*. This is due to the measure zero characteristic of each agent and our assumption that the society is at μ^*. A single agent who is assigned to play α_q^* instead of α_p^* then cannot affect the aggregate strategy level α^*. This is another way that the large population setting of our model simplifies our analysis. Given this payoff, the planner will be able to implement a particular transfer vector $t = (t_1, \cdots, t_n)$

[12]Nevertheless, we need the notation in (18) to capture the possibility of false reporting of type.

truthfully at μ^* if

$$\phi_p(p, \mu^*, t) \geq \phi_p(q, \mu^*, t) \tag{19}$$

for every $p, q \in \mathcal{P}$. In words, every agent should find it at least weakly preferable to play his own type specific strategy and receive his own type specific transfer than the strategy and transfer of any other type when society is at the efficient state. Therefore, these are the incentive compatibility (IC) constraints at μ^*, which we will generalize when we describe the mechanism more fully in Section 4.

The planner, therefore, seeks a transfer vector that minimizes the variance (17) while satisfying (19). Before characterizing the solution to this problem, let us consider two other alternatives which, as we will argue, cannot be that transfer scheme. First is the *equal redistribution transfer scheme*. Under this scheme, the planner redistributes an equal amount to every agent, irrespective of type. Budget balance then implies that every agent of every type p receives $t_p = T(\mu^*)$ as defined in (13). The resulting post redistribution payoffs (16) will then be strategically equivalent to $\hat{F}_{\alpha_p^*,p}(\mu^*)$ as defined in (12). But μ^* is the unique Nash equilibrium of $\hat{F}$ and α_p^* is the unique best response of every type p to μ^* (Proposition 2.4). Therefore, the equal redistribution rule will satisfy all IC constraints (19) and, in fact, will do so strictly. Hence, it may be possible to further improve equality by making truthful revelation weakly dominant and this is what we will discuss in Sections 3.1 and 4.

The fact that the equal redistribution scheme is not the solution to our problem is also the key difference between this paper and Lahkar and Mukherjee [11]. That paper sought to implement efficiency with budget balance in a large population public goods game. Externalities in a public goods model are positive. Hence, efficiency requires a subsidy for agents. Budget balance then involves a tax so that the revenue for the total subsidy is recovered. If this tax is equal for all agents, then that would be "equal redistribution" in that model. Such equal redistribution did indeed implement efficiency in that model and did so in strictly dominant strategies. But with equality being an additional objective in this model, we have to look beyond equal redistribution.

The second possibility is a transfer scheme that ensures *perfect equality*. This outcome would make the post redistribution payoff (16) of all agents perfectly equal. Thus, the planner would like to choose a transfer scheme $\tilde{t}$ such that at μ^*,

$$\alpha_p^* MP\left(\alpha^*\right) - c_p\left(\alpha_p^*\right) + \tilde{t}_p = \alpha_q^* MP\left(\alpha^*\right) - c_q\left(\alpha_q^*\right) + \tilde{t}_q \tag{20}$$

for all $p, q \in \mathcal{P}$. For a planner concerned with equality at the efficient state, this is obviously the *first best solution*. It is, however, easy to see that such a transfer scheme cannot satisfy incentive compatibility. Suppose $p < q$ so that, by Observation 2.2, $c_p(x) < c_q(x)$ for all $x \in \mathcal{S}$. Hence, $c_p(\alpha_q^*) < c_q(\alpha_q^*)$ which means

$$\begin{aligned} \alpha_q^* MP\left(\alpha^*\right) - c_p\left(\alpha_q^*\right) + \tilde{t}_q &> \alpha_q^* MP\left(\alpha^*\right) - c_q\left(\alpha_q^*\right) + \tilde{t}_q \\ &= \alpha_p^* MP\left(\alpha^*\right) - c_p\left(\alpha_p^*\right) + \tilde{t}_p, \end{aligned} \tag{21}$$

where the equality follows from (20). Thus, we have $\phi_p(q, \mu^*, \tilde{t}) > \phi_p(p, \mu^*, \tilde{t})$ so that type p's IC constraint (18) is violated. Hence, the first best solution cannot be achieved by the planner. We now discuss the *second best solution*.

3.1 Minimum Incentive Compatible Inequality at the Efficient State

Recall the variance (17), the budget balance condition (15) and the IC constraints (19). Formally, the planner's objective is to choose a transfer vector $t = (t_1, t_2, \cdots, t_n)$ so as to

$$\underset{t}{\text{Minimize}}\, V(\mu^*, t) \text{ such that } \phi_p(p, \mu^*, t) \geq \phi_p(q, \mu^*, t) \text{ and } \sum_p m_p t_p = T(\mu^*), \tag{22}$$

for all $p, q \in \mathcal{P}$. We characterize the solution to this problem through the following lemmas leading up to Proposition 3.4. All proofs are in Appendix A.2.

Lemma 3.1 *Recall the IC conditions (19). Consider a type $p \in \{1, 2, \cdots, n-1\}$ and an arbitrary transfer scheme $t = (t_1, t_2, \ldots, t_n)$ such that*

$$\phi_p(p, \mu^*, t) = \phi_p(p+1, \mu^*, t). \tag{23}$$

Then,

$$\phi_p(p, \mu^*, t) > \phi_p(p+q, \mu^*, t), \tag{24}$$

for all $q \in \{2, 3, \cdots, n-p\}$.

Lemma 3.1 implies that to ensure that agents do not claim to be of a higher type at μ^*, it suffices to equate their true payoff to the payoff they would obtain by claiming to be the next higher type. Suppose now that we have a transfer scheme $t = (t_1, \cdots, t_n)$ that satisfies Lemma 3.1. The following lemma then establishes certain characteristics of the payoffs resulting from those transfers as well as the transfers themselves.

Lemma 3.2 *Recall the payoff (18). Suppose the transfer scheme $t = (t_1, t_2, \cdots, t_n)$ satisfies (23) in Lemma 3.1. Then, the following hold.*

1. *$\phi_p(p, \mu^*, t) > \phi_{p+1}(p+1, \mu^*, t)$ for all $p = 1, 2, \ldots, n-1$.*

2. *$t_1 < t_2 < \ldots < t_n$.*

Lemma 3.2, therefore, establishes that under a transfer scheme that satisfies Lemma 3.1, types with a lower cost function obtain a higher payoff than types with a higher cost function. This is despite the fact, as part 2 of lemma shows, high cost types obtain a higher transfer. Part 2 of this lemma also leads to Lemma 3.3 that shows that agents will not have any incentive to claim to be of a lower type. Hence, Lemmas 3.1 and 3.3 suffice to rule out incentives for false representation.

Lemma 3.3 *Recall the IC conditions (19) and suppose a transfer scheme* $t = (t_1, t_2, \cdots, t_n)$ *satisfies part 2 of Lemma 3.2. Suppose* $p > q$*. Then,* $\phi_p(p, \mu^*, t) > \phi_p(q, \mu^*, t)$*. Therefore, if* t *satisfies (23), then all IC constraints (19) are satisfied.*

The key condition in Lemmas 3.1–3.3 is (23). The condition ensures there is no misrepresentation as a higher type in Lemma 3.1. It also gives rise to the ordering between the transfers in Lemma 3.2(2), which then leads to Lemma 3.3 that rules out misrepresentation as a lower type. Notice that the three preceding lemmas are independent of the budget balance condition. But once we combine the IC constraints with the budget balance condition, we obtain the unique solution to the planner's problem (22). The following proposition formalizes that solution. The proposition also shows that the solution satisfies *individual rationality*, which means the post redistribution payoffs will be positive for all types of agents. This is important because it means no agent has to be coerced to participate in the mechanism.

Proposition 3.4 *Consider the system of* n *linear equations consisting of the* $n-1$ *equations* $\phi_p(p, \mu^*, t) = \phi_p(p+1, \mu^*, t)$ *as specified in (23) for types* $p \in \{1, 2, \cdots, n-1\}$ *and the budget balance equation* $\sum_{p \in \mathcal{P}} m_p t_p = T(\mu^*)$*, where* $T(\mu^*)$ *is as defined in (13). Denote the solution to these* n *equations as* $t^* = (t_1^*, t_2^*, \cdots, t_n^*)$*. Then,* t^* *is the solution to the planner's problem (22).*

Thus, t^* *satisfies* $t_1^* < t_2^* < \cdots < t_n^*$*. Moreover, among all transfer vectors* t *at the efficient state that satisfy incentive compatibility and budget balance,* t^* *maximizes the post redistribution payoff of type* n *agents,* $\phi_n(n, \mu^*, t)$*. Hence, at* t^**,*

$$\phi_1(1, \mu^*, t^*) > \phi_2(2, \mu^*, t^*) > \cdots > \phi_n(n, \mu^*, t^*) > 0, \tag{25}$$

which means t^* *also ensures individual rationality.*

Proposition 3.4 is the most important technical result of our paper. It characterizes the optimal transfer vector $t^* = (t_1^*, t_2^*, \cdots, t_n^*)$ as the solution to a set of linear equations. We emphasise that this proposition is not an implementation result. We will present our main theorem on dominant strategy implementation of (μ^*, t^*) in the next section using Proposition 3.4.

Even though t^* does not ensure perfect equality, it does minimize inequality at the efficient state subject to incentive compatibility and budget balance. Any other transfer vector that satisfies these two conditions must generate a higher variance (17) in the post redistribution payoffs. One such transfer scheme is the equal redistribution scheme $T(\mu^*)$ as defined in (13). Compared to equal redistribution, t^* increases the payoff of higher cost types and reduces the payoff of lower cost types thereby reducing variance while still satisfying incentive compatibility. Quite intuitively, and as implied by Lemma 3.2(2), the variance minimizing transfer favors high cost agents over low cost ones. Hence, $t_1^* < t_2^* < \cdots < t_n^*$.

The order among the payoffs in (25) arises from part 1 of Lemma 3.2. All these payoffs are strictly positive because the equal redistribution transfer $T(\mu^*)$ itself ensures that the payoff of the highest cost type n is strictly positive.[13] But the proof of Proposition 2.4 not only shows that t^* minimizes the variance but also maximizes the payoff of type n agents among all incentive compatible and budget balanced transfer vectors.[14] Hence, the payoff of type n agents must be even higher than under equal redistribution. Therefore, not only do we minimize variance at the efficient state but also maximizes the minimum payoff. In this sense, implementing (μ^*, t^*) would be one way to reconcile the utilitarian objective of achieving efficiency with the Rawlsian objective of maximizing the minimum payoff. However, as we discuss in more detail in Section 5, this is not equivalent to implementing the Rawlsian social choice function.

We have interpreted our results as the planner imposing the externality equivalent tax $\alpha_p^*(AP(\alpha^*) - MP(\alpha^*))$ on agents of type p and providing them the transfer t_p^*. It is worth pointing out that we could also have provided an alternative but equivalent presentation in terms of the optimal payment vector and not the optimal transfer vector t^*. Suppose β_p is the (net) payment received by a type p agent. Then the payment vector $\beta^* = (\beta_1^*, \beta_2^*, \cdots, \beta_n^*)$ such that

$$\beta_p^* = \alpha_p^*(MP(\alpha^*) - AP(\alpha^*)) + t_p^* \tag{26}$$

would solve the planner's problem (22) subject to the budget constraint $\sum_{p \in \mathcal{P}} m_p \beta_p = 0$.[15] In this interpretation, these optimal payments would be uniquely defined but not the taxes and transfers. In our presentation, though, because we have fixed the tax at $\alpha_p^*(AP(\alpha^*) - MP(\alpha^*))$, we also obtain a unique value of t_p^*. According to us, it is economically more intuitive to present our conclusions in this manner. This tax is equal to the negative externality an agent is generating at the efficient state. Hence, by imposing this particular tax, it is as if the planner is using the tax to achieve efficiency and then using the transfers as a redistributive measure while retaining efficiency.

4 Dominant Strategy Implementation

Proposition 3.4 characterizes the transfer vector t^* that minimizes inequality at the efficient state μ^* while retaining incentive compatibility and budget balance. It, however, required the assumption that the planner knows the type distribution m. To make our problem more substantive, we now drop this assumption as well as the assumption that the society is at μ^*. This section then describes the mechanism that enables the planner to simultaneously implement both μ^* and t^* in dominant strategies. Formally, for any given type distribution m, the planner wishes to implement μ^* and t^* corresponding to that type distribution. Thus, $m \mapsto (\mu^*, t^*)$ is the planner's social choice function. While the planner doesn't know the type distribution, we do assume that he knows the set of

[13] Even without $T(\mu^*)$, the fact that there are no fixed costs in our model ensures that the pre–redistribution payoff $\alpha_n^* MP(\alpha^*) - c_n(\alpha_n^*)$ in (16) is strictly positive.

[14] See Claim 5 of that proof.

[15] With an abuse of notation and by using the original tragedy of the commons payoff (2), we can then write the analog of payoff (18) as $\phi_p(q, \mu^*, \beta) = \alpha_q^* AP(\alpha^*) - c_p(\alpha_q^*) + \beta_q$. The IC constraints (19) will then take the form $\phi_p(p, \mu^*, \beta) \geq \phi_p(q, \mu^*, \beta)$ for all $p, q \in \mathcal{P}$ and the variance (17) would then have to be written as $V(\mu^*, \beta) = \sum_{p \in \mathcal{P}} m_p \left[\alpha_p^* AP\left(\alpha^*\right) - c_p\left(\alpha_p^*\right) + \beta_p - \bar{F}(\mu^*)\right]^2$.

possible types $\mathcal{P}$, the production function π and the type specific cost functions $(c_1, \cdots, c_n)$.

By the revelation principle, it suffices to consider direct mechanisms. Hence, the planner designs a direct mechanism, which we denote as Φ, as follows. The planner asks each agent to report his type. Suppose $\tilde{m} = (\tilde{m}_1, \tilde{m}_2, \cdots, \tilde{m}_n)$ is the reported type distribution. Thus, $\tilde{m}_p$ is the proportion of agents who report their type to be p. As agents can report type falsely, it is possible that $\tilde{m}_p \neq m_p$. Using the reported type distribution $\tilde{m}$, the planner calculates the efficient state corresponding to $\tilde{m}$. This can be done by proceeding as in Proposition 2.4 once m is replaced with $\tilde{m}$ in (10). Let the efficient state corresponding to the distribution $\tilde{m}$ be $\tilde{\mu}^*$ and the strategy level of a type p agent at that state be $\tilde{\alpha}_p^*$. Thus, $\tilde{\mu}^* = \left(\tilde{m}_1 \delta_{\tilde{\alpha}_1^*}, \tilde{m}_2 \delta_{\tilde{\alpha}_2^*}, \cdots, \tilde{m}_n \delta_{\tilde{\alpha}_n^*}\right)$. Denote the corresponding aggregate strategy level $A(\tilde{\mu}^*) = \sum_{p \in \mathcal{P}} \tilde{m}_p \tilde{\alpha}_p^* = \tilde{\alpha}^*$. Further, analogous to (13) and (18), we define

$$T(\tilde{\mu}^*) = \tilde{\alpha}^* \left(AP(\tilde{\alpha}^*) - MP(\tilde{\alpha}^*)\right). \tag{27}$$

and

$$\phi_p(q, \tilde{\mu}^*, \tilde{t}) = \tilde{\alpha}_q^* MP(\tilde{\alpha}^*) - c_p(\tilde{\alpha}_q^*) + \tilde{t}_q. \tag{28}$$

for some arbitrary transfer vector $\tilde{t} = (\tilde{t}_1, \tilde{t}_2, \cdots, \tilde{t}_n)$.

Intuitively, (27) is the aggregate tax that the planner would receive if all agents who report type to be q are assigned and play $\tilde{\alpha}_q^*$ while (28) is the payoff of a type p agent who reports type to be q and is, therefore, assigned strategy $\tilde{\alpha}_q^*$ and transfer $\tilde{t}_q$. Following Proposition 3.4, we now denote as $\tilde{t}^* = (\tilde{t}_1^*, \tilde{t}_2^*, \cdots, \tilde{t}_n^*)$ the solution to the following system of equations

$$\begin{aligned} &\phi_p(p, \tilde{\mu}^*, \tilde{t}) = \phi_p(p+1, \tilde{\mu}^*, \tilde{t}), \text{ for all } p \in \{1, 2, \cdots, n-1\} \\ &\sum_{p \in \mathcal{P}} \tilde{m}_p t_p = T(\tilde{\mu}^*), \end{aligned} \tag{29}$$

where $T(\tilde{\mu}^*)$ is as defined in (27). The first set of equalities are the minimal IC constraints that need to satisfied as in Lemma 3.1 but with respect to $\tilde{m}$. The second equality is the budget balance condition similar to (15) but with respect to $\tilde{m}$.

The planner then assigns the type specific strategy $\tilde{\alpha}_q^*$ and the transfer $\tilde{t}_q^*$ to any agent who announces type to be q. Thus, in the conventional terminology of mechanism design, the planner designs the direct mechanism

$$\Phi : (q, \tilde{m}) \mapsto (\tilde{\alpha}_q^*, \tilde{t}_q^*) \tag{30}$$

which takes the reported type q of an agent and the reported type distribution $\tilde{m}$ as inputs and generates the type specific strategy and transfer $(\tilde{\alpha}_q^*, \tilde{t}_q^*)$ as output as described above. The resulting payoff is then $\phi_p(q, \tilde{\mu}^*, \tilde{t}^*)$ as defined in (28). The following is then the main result of this paper.

Theorem 4.1 *The direct mechanism Φ defined by (30) implements (μ^*, t^*) in weakly dominant strategies, where μ^* is the efficient state characterized in Proposition 2.4 and t^* is the transfer vector characterized in Proposition 3.4. The mechanism also satisfies budget balance and individual* *rationality. The resulting variance in payoffs is*

$$V\left(\mu^{*},t^{*}\right)=\sum_{p\in\mathcal{P}}m_{p}\left[\alpha_{p}^{*}MP\left(\alpha^{*}\right)-c_{p}\left(\alpha_{p}^{*}\right)+t_{p}^{*}-\bar{F}(\mu^{*})\right]^{2}, \tag{31}$$

which is the lowest possible variance at the efficient state subject to incentive compatibility and budget balance.

Proof. A single agent cannot influence the type distribution $\tilde{m}$ and, hence, the aggregate strategy level $\tilde{\alpha}^{*}$ or the aggregate tax $T(\tilde{\mu}^{*})$. Consider an agent p. Given $\tilde{\alpha}^{*}$, $(\tilde{\alpha}_{p}^{*},\tilde{t}_{p}^{*})$ satisfy (29). We now apply arguments akin to Lemmas 3.1–3.3 and Proposition 3.4 but with respect to the reported type distribution $\tilde{m}$. For all $\tilde{m}$, it is weakly incentive compatible for type p to reveal type truthfully and, if fact, strictly so if $p=n$. Hence, $\tilde{m}=m$ and $(\tilde{\mu}^{*},\tilde{t}^{*})=(\mu^{*},t^{*})$ gets implemented. The conclusions about budget balance and individual rationality follow from Proposition 3.4. The resulting variance (31) follows from (17). ■

The logic behind Theorem 4.1 is the same as that of Proposition 3.4, which established dominance of truthful revelation at (μ^{*},t^{*}) at the actual type distribution m. But mathematically, there is noting special about m. Hence, if the planner announces that he will calculate $(\tilde{\mu}^{*},\tilde{t}^{*})$ and assign strategies and transfers based on the reported type, it becomes a dominant strategy for every agent to report type truthfully. The type distribution that gets revealed is the true one m and, therefore, the outcome that is implemented is (μ^{*},t^{*}).

As we noted after Proposition 3.4, t^{*} is different from the equal redistribution transfer scheme. It also achieves a lower inequality at μ^{*}. It would be instructive to understand why this happens. While discussing equal redistribution, we noted that such a transfer scheme will render truthful revelation strictly dominant. This creates the possibility of adjusting the equal redistribution transfer scheme such that at least some of those constraints are satisfied with equality, which would render truthful revelation only weakly dominant for at least some types. It turns out we can do for all types from 1 to $n-1$, as argued in Proposition 3.4. This allows us to reduce the transfers for the lower cost agents and increase them for the higher cost agents in comparison to equal redistribution. Then, while still satisfying incentive compatibility and budget balance, we are able to reduce inequality at μ^{*}.

We have noted earlier that this paper does not rely on the evolutionary implementation methodology applied in large population games by Sandholm [17, 18, 19] and Lahkar and Mukherjee [10, 12]. It would be useful to remark why. Evolutionary implementation relies on the fact that the externality adjusted game $\hat{F}$ defined in (8) is a potential game. Standard evolutionary dynamics would, therefore, converge to its Nash equilibrium which, of course, is the efficient state of F. But once we add transfers to $\hat{F}$ with the objective of reducing inequality, we no longer obtain the potential game property. The method of evolutionary implementation is then not applicable, which necessitates the more conventional dominant strategy implementation approach. As noted earlier, such a dominant strategy implementation approach has also been applied by Lahkar and Mukherjee [11] in their large population public goods model. If, similar to that paper, our objective had also been restricted to implementing just μ^*, we could have done so using the equal redistribution transfer $T(\tilde{\mu}^*)$. But because we are also concerned with inequality, we have to suitably modify our dominant strategy implementation approach to also characterize and implement the inequality minimizing transfer vector t^*.

5 Rawlsian Outcome

Our focus in this paper has been on minimizing inequality at efficiency subject to incentive compatibility and budget balance (Theorem 4.1). As a byproduct of minimizing inequality, Proposition 3.4 also shows that our solution maximizes the lowest payoff at efficiency, again subject to incentive compatibility and budget balance. This suggests a connection between our problem and the problem of implementing the Rawlsian social choice function. The Rawlsian social choice function would seek to implement an outcome that maximizes the lowest payoff. In our model, if incentive compatibility is not a concern, then the Rawlsian outcome is simply the efficient state μ^* and the transfer vector that ensures perfect equality.[16] But as we argued, such perfect equality at efficiency is not incentive compatible. Instead, by Proposition 3.4, if we restrict ourselves to the efficient state μ^*, then the best feasible solution to the Rawlsian problem is the transfer vector t^*.

But suppose we are willing to sacrifice efficiency while still requiring incentive compatibility and budget balance. Then, can we find an outcome (μ, t), where $\mu \neq \mu^*$ is a social state in F and t is a transfer vector satisfying budget balance and incentive compatibility, that ensures a higher "minimum payoff" than (μ^*, t^*)? If so, then that will imply the Rawlsian outcome over all possible social states will be different from (μ^*, t^*). We explore this question in this section. We do not attempt a general characterization of the solution. Instead, we provide a numerical example that shows that the Rawlsian outcome is not necessarily (μ^*, t^*). The example will also allow us to illustrate our characterization of (μ^*, t^*).

Example 5.1 *Consider the model described in Section 2 with strategy set $\mathcal{S} = (0, \infty)$. Let the set of populations or types be $\mathcal{P} = \{1, 2, 3\}$ and the type distribution be $(m_1, m_2, m_3) = (0.2, 0.3, 0.5)$. Suppose the type specific cost functions are $c_p(x) = k_p x^2$ where $\{k_1, k_2, k_3\} = \{1, 2, 3\}$ and the production function is $\pi(\alpha) = 10\sqrt{\alpha}$. The average product is, therefore, $\frac{10\sqrt{\alpha}}{\alpha} = \frac{10}{\sqrt{\alpha}}$. Hence, given a social state μ with aggregate strategy $A(\mu) = \alpha$, the payoff of a type $p \in \{1, 2, 3\}$ agent in the tragedy of the commons F defined by (2) is*

$$F_{x,p}(\mu) = \frac{10x}{\sqrt{\alpha}} - k_p x^2. \tag{32}$$

Applying Proposition 2.3, we can characterize the Nash equilibrium μ^N of Example 5.1. The Nash equilibrium involves type p agents playing strategy α_p^N with $(\alpha_1^N, \alpha_2^N, \alpha_3^N) = (3.644, 1.822, 1.215)$.

[16] See the discussion preceding Section 3.1.

The marginal product is $\pi'(\alpha) = \frac{5}{\sqrt{\alpha}}$. Therefore, the externality (7) an agent playing strategy x generates is $\frac{-5x}{\sqrt{\alpha}}$. Hence, from (32), we obtain the externality adjusted payoff (8) in our example,

$$\hat{F}_{x,p}(\mu) = \frac{5x}{\sqrt{\alpha}} - k_p x^2. \tag{33}$$

Proposition 2.4 then yields the efficient state of Example 5.1, which is $\mu^* = (m_1\delta_{\alpha_1^*}, m_2\delta_{\alpha_2^*}, m_3\delta_{\alpha_3^*})$ where

$$(\alpha_1^*, \alpha_2^*, \alpha_3^*) = (2.2956, 1.1478, 0.7652). \tag{34}$$

The aggregate strategy level at the efficient state is $\alpha^* = \sum_{p\in\mathcal{P}} m_p\alpha_p^* = 1.186$. The efficient state, therefore, involves every agent of type p paying a tax $\frac{5\alpha_p^*}{\sqrt{\alpha^*}}$ so that the total tax revenue (13) raised by the planner is $T(\mu^*) = \frac{5\sum_p m_p\alpha_p^*}{\sqrt{\alpha^*}} = 5\sqrt{\alpha^*} = 5.4453$.

We now introduce transfer vectors $t = (t_1, t_2, t_3)$. To characterize the transfer vector t^* that minimizes inequality at the efficient state μ^*, we write the payoff (18) of a type p agent who reports type to be q at μ^* in Example 5.1 as

$$\begin{aligned}
\phi_p(q, \mu^*, t) &= \alpha_q^* MP(\alpha^*) - k_p(\alpha_q^*)^2 + t_q \\
&= \frac{5\alpha_q^*}{\sqrt{1.186}} - k_p(\alpha_q^*)^2 + t_q \\
&= 4.5911\alpha_q^* - k_p(\alpha_q^*)^2 + t_q.
\end{aligned} \tag{35}$$

Notice that in Example 5.1, $k_1 < k_2 < k_3$. Hence, by Proposition 3.4, the binding IC (incentive compatibility) constraints for t^* are $\phi_1(1, \mu^*, t_1) = \phi_1(2, \mu^*, t_2)$ and $\phi_2(2, \mu^*, t_2) = \phi_2(3, \mu^*, t_3)$. In addition, we have the budget balance condition $\sum_{p\in\mathcal{P}} m_p t_p = T(\mu^*)$, which we have previously calculated to be 5.4453. By (35) and the values of m_p, k_p in Example 5.1, these constraints take the form

$$4.5911\alpha_1^* - (\alpha_1^*)^2 + t_1 = 4.5911\alpha_2^* - (\alpha_2^*)^2 + t_2, \tag{36}$$

$$4.5911\alpha_2^* - 2(\alpha_2^*)^2 + t_2 = 4.5911\alpha_3^* - 2(\alpha_3^*)^2 + t_3, \tag{37}$$

$$0.2t_1 + 0.3t_2 + 0.5t_3 = 5.4453, \tag{38}$$

with (36) and (37) being the IC constraints for types 1 and 2 respectively, and (38) being the budget balance condition. Solving these equations, we obtain the desired transfer vector t^* to be

$$(t_1^*, t_2^*, t_3^*) = (4.245, 5.5624, 5.8551). \tag{39}$$

Thus, (μ^*, t^*) as characterized in (34) and (39) implements the planner's objective of achieving efficiency with minimum inequality consistent with incentive compatibility and budget balance. By

(35), the resulting payoffs at the efficient state following redistribution for the three types are

$$(\phi_1(1, \mu^*, t^*), \phi_2(2, \mu^*, t^*), \phi_3(3, \mu^*, t^*)) = (9.5146, 8.1972, 7.6117). \tag{40}$$

The aggregate payoff of agents at this outcome is $\sum_{p=1}^{n} m_p \phi_1(1, \mu^*, t^*) = 8.1679$ which, by definition of the efficient state μ^*, is the highest possible in Example 5.1. As implied by Proposition 3.4, type 3 agents, who have the highest cost of effort, have the lowest payoff in (40).

We now consider the Rawlsian social choice function. Thus, instead of minimizing inequality at the efficient state, the planner wishes to implement an outcome that maximizes the lowest payoff across all social states in F. To explore this possibility, it suffices to consider states in monomorphic population states. This would ensure that within each population at least, payoffs would be equal. Thus, let $\mu = (m_1\delta_{\alpha_1}, m_2\delta_{\alpha_2}, m_3\delta_{\alpha_3})$ be such a social state in Example 5.1. The aggregate strategy is, therefore, $\alpha = \sum_{p=1}^{3} m_p \alpha_p$. Further, let $t = (t_1, t_2, t_3)$ be a transfer vector. Analogous to (18), denote as

$$\begin{aligned} \phi_p(q, \mu, t) &= \alpha_q AP(\alpha) - c_p(\alpha_q) + \alpha_q(MP(\alpha) - AP(\alpha)) + t_q \\ &= \alpha_q MP(\alpha) - c_p(\alpha_q) + t_q \\ &= \frac{5\alpha_q}{\sqrt{\alpha}} - k_p \alpha_q^2 + t_q, \end{aligned} \tag{41}$$

the post redistribution payoff of a type p agent who reports type to be q at the social state μ and the transfer vector t. Unlike (35), (41) is defined at all social states of the form $\mu = (m_1\delta_{\alpha_1}, m_2\delta_{\alpha_2}, m_3\delta_{\alpha_3})$. Thus, an agent claiming to of type q is assigned the strategy and transfer (α_q, t_q), pays the tax $\alpha_q(AP(\alpha) - MP(\alpha))$ and plays the tragedy of the commons (32). We will focus on the incentive compatible solution where every agent will reveal type truthfully. Hence, analogous to (13), the aggregate tax paid by the agents in (41) would be

$$T(\mu) = \sum_{p=1}^{3} m_p \alpha_p (AP(\alpha) - MP(\alpha)) = 5\sqrt{\alpha} = 5\sqrt{0.2\alpha_1 + 0.3\alpha_2 + 0.5\alpha_3}. \tag{42}$$

Maximizing the minimum payoff means that for every possible m, the planner wishes to implement the Rawlsian social choice function $m \mapsto (\mu^R, t^R)$ where (μ^R, t^R), called the Rawlsian outcome, solves

$$\begin{aligned} &\max_{(\mu,t)} [\min\{\phi_1(1, \mu, t), \phi_2(2, \mu, t), \phi_3(3, \mu, t)\}] \\ &\quad \text{subject to } \mu = (m_1\delta_{\alpha_1}, m_2\delta_{\alpha_2}, m_3\delta_{\alpha_3}) \\ &\quad \phi_p(p, \mu, t) \geq \phi_p(q, \mu, t) \text{ for all } p, q \in \mathcal{P} \\ &\quad m_1 t_1 + m_2 t_2 + m_3 t_3 = 5\sqrt{0.2\alpha_1 + 0.3\alpha_2 + 0.5\alpha_3}. \end{aligned} \tag{43}$$

The constraints $\phi_p(p, \mu, t) \geq \phi_p(q, \mu, t)$ are the IC constraints, with $\phi_p(q, \mu, t)$ being as defined in

(41). The last constraint in (43) is the budget balance condition arising from (42).

Through arguments similar to those leading up to Proposition 3.4, we can show that to satisfy the IC constraints, it suffices to make $\phi_p(p, \mu, t) = \phi_p(p+1, \mu, t)$ for all $p \in \mathcal{P}$. To characterize (μ^*, t^*), these equalities along with the budget balance condition sufficed, as can be seen from (36)–(38). But in the present scenario where $(\alpha_1, \alpha_2, \alpha_3)$ are unknown, these four equations by themselves will not be enough. Instead, we need to explicitly solve the maximization exercise (43). The numerical solution that we obtain is

$$\left(\alpha_1^R, \alpha_2^R, \alpha_3^R\right) = (2.4264, 0.9099, 0.6066), \tag{44}$$

$$\left(t_1^R, t_2^R, t_3^R\right) = (3.0358, 5.3356, 5.8876). \tag{45}$$

Thus, the social state at the Rawlsian outcome is $\mu^R = (m_1\delta_{\alpha_1^R}, m_2\delta_{\alpha_2^R}, m_3\delta_{\alpha_3^R})$. The aggregate strategy at this outcome is $\alpha^R = \sum_{p\in\mathcal{P}} m_p\alpha_p^R = 1.0616$. Inserting (44) and (45) in (41) and using $\{k_1, k_2, k_3\} = \{1, 2, 3\}$ from Example 5.1, we obtain the type specific payoffs at the Rawlsian outcome to be

$$\left(\phi_1(1, \mu^R, t^R), \phi_2(2, \mu^R, t^R), \phi_3(3, \mu^R, t^R)\right) = (8.9233, 8.0953, 7.7274). \tag{46}$$

The calculation in (46) is based on the fact that due to incentive compatibility, all agents reveal their type truthfully. Therefore, the payoff of type p agents is $\phi_p(p, \mu^R, t^R)$.

Like in (40) and as is to expected given the cost disadvantage, type 3 agents have the lowest payoff in (46). But their payoff is higher in (46). Therefore, the Rawlsian outcome (μ^R, t^R) achieves a higher minimum payoff in the society than the outcome (μ^*, t^*) that minimizes inequality at the efficient state. As is also evident from (34) and (44), $\mu^* \neq \mu^R$. Hence, implementing the Rawlsian social choice function involves a sacrifice of efficiency. This can also be seen from fact that the aggregate payoff in (46), $\sum_{p=1}^{3} m_p\phi_p(p, \mu^R, t^R) = 8.0769 < 8.1679$, the aggregate payoff at the efficient state. Thus, Example 5.1, while not providing a general characterization of the Rawlsian outcome, does show that subject to incentive compatibility and budget balance, implementing efficiency with minimum inequality is not equivalent to implementing the Rawlsian social choice function.

We have not provided any detailed description of the mechanism the planner can use to implement (μ^R, t^R). Briefly, however, the mechanism would largely be as in Section 4. The planner asks for reports of type. Suppose the reported distribution is $\tilde{m}$. The planner calculates the corresponding Rawlsian outcome $(\tilde{\mu}^R, \tilde{t}^R)$ by maximizing (43) but with respect to the distribution $\tilde{m}$. He then assigns the corresponding strategy $(\tilde{\alpha}_q^R, \tilde{t}_q^R)$ to any agent who reports type q. Arguments akin to Proposition 3.4 and Theorem 4.1 will then imply that truthful revelation is dominant for all agents. Finally, we point out that just as the optimal payment (26) that implements efficiency with minimum inequality, we could also have presented our calculation of the Rawlsian outcome in terms of an optimal payment vector $\beta^R = (\beta_1^R, \beta_2^R, \cdots, \beta_n^R)$ where $\beta_p^R = \alpha_p^R(MP(\alpha^R) - AP(\alpha^R)) + t_p^R$. But fixing the tax at the externality level $\alpha_p^R(AP(\alpha^R) - MP(\alpha^R))$ and then characterizing the unique t_p^R allows for a more direct comparison with our approach to implementing efficiency with minimum inequality. It is important to note that such algebraic manipulations will have no affect on the Rawlsian social state μ^R or the Rawlsian payoffs (46), which are what we are most interested in this section.

6 Conclusion

We have considered the implementation of efficiency with minimum inequality in a large population model of negative externalities. Agents are of different types which are distinguished by cost functions that are private information. Total output is a function of aggregate strategy which is shared among agents according to individual strategy. The model is, therefore, equivalent to a tragedy of the commons. Imposition of externality equivalent taxes restores efficiency in the model. The planner would like to redistribute the tax revenue as transfers so as to reduce inequality, as measured by the variance of payoffs, at the efficient state while being subject to incentive compatibility and budget balance.

We first characterize the inequality minimizing vector of type specific transfers. We then describe a mechanism that would enable the planner to implement both the efficient state and the inequality minimizing transfer vector in dominant strategies. The planner asks agents to report their types and calculates the efficient state, externality tax and inequality minimizing transfers based on reported types. Due to the large population characteristic of the model, it then becomes weakly dominant for all agents to report the type truthfully thereby implementing the desired objective of the planner. Finally, while minimizing inequality at efficiency also ensures maximization of minimum payoff at the efficient state, it is not equivalent to implementing the Rawlsian social choice function. There may exist other states which are not efficient but where, through appropriate transfers, it is possible to further improve the welfare of the most disadvantaged agents in an incentive compatible manner while satisfying budget balance.

An important research question that arises is a more general characterization of the Rawlsian social state in a large population model. In the present paper, we have only provided a counterexample because that suffices to show that the inequality minimizing efficient outcome is not the Rawlsian outcome. But independent of efficiency, the Rawlsian outcome is interesting on its own and a more rigorous analysis of this outcome is worth exploring. We have also only sketched out the mechanism for implementing the Rawlsian outcome in the last paragraph of Section 5. The question of establishing the details of this mechanism remain. Generalizing the present analysis of efficiency with minimum inequality to models other than the tragedy of the commons also remains an unexplored question.

A Appendix

A.1 Appendix to Section 2

Proof of Proposition 2.3: In the discussion preceding Proposition 2.3, we have argued that $b_p(\alpha)$ is the unique best response of type p agents to any state μ such that $A(\mu) = \alpha$. The additional assumption that $c_p'(0) = 0$ then implies that for every $p \in \mathcal{P}$, this unique best response in (2) satisfies

$$AP(\alpha) = c_p'(b_p(\alpha)), \tag{47}$$

with $b_p(\alpha) \in (0, \infty)$ for all $\alpha \in (0, \infty)$.

Proposition 3.1 in Lahkar [9] shows that in large population aggregative games such as (2), all Nash equilibria can be characterized as solutions to (3). Due to our assumptions that $AP(\alpha)$ is strictly declining and c_p is strictly convex, we conclude from (47) that $b_p(\alpha)$ is strictly declining for all p. Hence, (3) has a unique solution, which we denote as α^N. By Proposition 3.1 in Lahkar [9], we then obtain the unique Nash equilibrium μ^N as defined in (4) where all type p agents play $b_p(\alpha^N)$. The aggregate strategy level at μ^N is, therefore, $\alpha^N = \sum_{p\in\mathcal{P}} m_p \alpha_p^N$ and condition (5) follows from (47). ■

Proof of Proposition 2.4: We first establish that (10) has a unique solution. The assumptions of our model imply that the unique best response $\hat{b}_p(\alpha)$ in $\hat{F}$ is characterized by

$$MP(\alpha) = c_p'\left(\hat{b}_p(\alpha)\right). \tag{48}$$

Due to the strict concavity of π, $MP(\alpha)$ is strictly declining. Hence, by a similar argument as in Proposition 2.3, $\hat{b}_p(\alpha)$ is strictly declining. This establishes uniqueness of the solution to (10). The argument in the proof of Proposition 2.3 then implies that μ^* is the unique Nash equilibrium of $\hat{F}$. The remaining conclusions follow from the discussion preceding Proposition 2.4. Proposition 5.6 in Lahkar and Mukherjee [11] shows this Nash equilibrium of $\hat{F}$ is also the efficient state of F defined by (2). The conclusion $\alpha_p^* < \alpha_p^N$ follows from (5), (11), the strict convexity of c_p and the fact that $MP(\alpha) < AP(\alpha)$. Condition (11) follows from (48). ■

Proof of Corollary 2.5: Since μ^* is the unique Nash equilibrium of the game $\hat{F}$ characterized by (8) and every agent has a unique best response to every state, α_p^* is the unique best response to μ^* for a type p agent. Therefore,

$$\alpha_p^* MP(\alpha^*) - c_p(\alpha_p^*) > \alpha_q^* MP(\alpha^*) - c_p(\alpha_q^*). \tag{49}$$

By Observation 2.2, as $p < q$, $c_p(\alpha_q^*) < c_q(\alpha_q^*)$. Therefore,

$$\alpha_q^* MP(\alpha^*) - c_p(\alpha_q^*) > \alpha_q^* MP(\alpha^*) - c_q(\alpha_q^*). \tag{50}$$

Combining (49) and (50) and using (12), we obtain $\hat{F}_{\alpha_p^*,p}(\alpha^*) > \hat{F}_{\alpha_q^*,q}(\alpha^*)$. For the relationship between α_p^* and α_q^*, note from (11) that $c_p'(\alpha_p^*) = c_q'(\alpha_q^*) = MP(\alpha^*)$. By Assumption 2.1, if $p < q$, then $c_p'(\alpha_p^*) < c_q'(\alpha_p^*)$. The strict convexity of cost functions then imply that $\alpha_p^* > \alpha_q^*$. ■

A.2 Appendix to Section 3

Proof of Lemma 3.1: We start with type $n-1$ and proceed in reverse order. For type $n-1$, (24) doesn't apply but suppose (23) holds. Thus,

$$\begin{aligned}&\phi_{n-1}(n-1, \mu^*, t) = \phi_{n-1}(n, \mu^*, t) \\ \Rightarrow &\alpha_{n-1}^* MP(\alpha^*) - c_{n-1}(\alpha_{n-1}^*) + t_{n-1} = \alpha_n^* MP(\alpha^*) - c_{n-1}(\alpha_n^*) + t_n. \end{aligned} \tag{51}$$

Now consider type $n-2$. Suppose (23) holds. Hence, $\phi_{n-2}(n-2, \mu^*, t) = \phi_{n-2}(n-1, \mu^*, t)$ or

$$\alpha_{n-2}^* MP(\alpha^*) - c_{n-2}(\alpha_{n-2}^*) + t_{n-2} = \alpha_{n-1}^* MP(\alpha^*) - c_{n-2}(\alpha_{n-1}^*) + t_{n-1}. \tag{52}$$

Then, to show (24), we need to show $\phi_{n-2}(n-2, \mu^*, t) > \phi_{n-2}(n, \mu^*, t)$. For this, we can use (52) and show $\phi_{n-2}(n-1, \mu^*) > \phi_{n-2}(n, \mu^*)$ or

$$\alpha_{n-1}^* MP(\alpha^*) - c_{n-2}(\alpha_{n-1}^*) + t_{n-1} > \alpha_n^* MP(\alpha^*) - c_{n-2}(\alpha_n^*) + t_n \tag{53}$$

Notice that we can derive (53) from (51) by adding $c_{n-1}(\alpha_{n-1}^*) - c_{n-2}(\alpha_{n-1}^*)$ and $c_{n-1}(\alpha_n^*) - c_{n-2}(\alpha_n^*)$ to the LHS and RHS of (51) respectively. But $\alpha_{n-1}^* > \alpha_n^*$ (Corollary 2.5). Hence, Assumption 2.1, $c_{n-1}(\alpha_{n-1}^*) - c_{n-2}(\alpha_{n-1}^*) > c_{n-1}(\alpha_n^*) - c_{n-2}(\alpha_n^*)$. But then, this establishes (53).

For type $n-3$, we need to argue that if $\phi_{n-3}(n-3, \mu^*, t) = \phi_{n-3}(n-2, \mu^*, t)$, then $\phi_{n-3}(n-3, \mu^*, t) > \phi_{n-3}(n-1, \mu^*, t)$ and $\phi_{n-3}(n-3, \mu^*, t) > \phi_{n-3}(n, \mu^*, t)$. To show $\phi_{n-3}(n-3, \mu^*, t) > \phi_{n-3}(n-1, \mu^*, t)$, we proceed as before for type $n-2$ and show $\phi_{n-3}(n-2, \mu^*, t) > \phi_{n-3}(n-1, \mu^*, t)$. This follows from (52) if we add $c_{n-2}(\alpha_{n-2}^*) - c_{n-3}(\alpha_{n-2}^*)$ and $c_{n-2}(\alpha_{n-1}^*) - c_{n-3}(\alpha_{n-1}^*)$ to the LHS and RHS of (52) respectively and then note that because $\alpha_{n-2}^* > \alpha_{n-1}^*$, $c_{n-2}(\alpha_{n-2}^*) - c_{n-3}(\alpha_{n-2}^*) > c_{n-2}(\alpha_{n-1}^*) - c_{n-3}(\alpha_{n-1}^*)$.

To show $\phi_{n-3}(n-3, \mu^*, t) > \phi_{n-3}(n, \mu^*, t)$, the above argument means it suffices to show $\phi_{n-3}(n-1, \mu^*, t) > \phi_{n-3}(n, \mu^*, t)$. For that, we use (53), which we have established. Note that $\phi_{n-3}(n-1, \mu^*, t) > \phi_{n-3}(n, \mu^*, t)$ is equivalent to

$$\alpha_{n-1}^* MP(\alpha^*) - c_{n-3}(\alpha_{n-1}^*) + t_{n-1} > \alpha_n^* MP(\alpha^*) - c_{n-3}(\alpha_n^*) + t_n \tag{54}$$

We can obtain (54) by adding $c_{n-2}(\alpha_{n-1}^*) - c_{n-3}(\alpha_{n-1}^*)$ and $c_{n-2}(\alpha_n^*) - c_{n-3}(\alpha_n^*)$ to the LHS and RHS of (53) respectively. The desired conclusion then follows by noting that because $\alpha_{n-1}^* > \alpha_n^*$, $c_{n-2}(\alpha_{n-1}^*) - c_{n-3}(\alpha_{n-1}^*) > c_{n-2}(\alpha_n^*) - c_{n-3}(\alpha_n^*)$.

For the remaining types $p \in \{1, 2, \cdots, n-4\}$, we can proceed similarly through an inductive argument. For each type p, we use the arguments established for type $p+1$ and prove the claim. This would establish the lemma. ■

Proof of Lemma 3.2: By (23), we have $\phi_p(p, \mu^*, t) = \phi_p(p+1, \mu^*, t)$ for all $p \in \{1, 2, \cdots, n-1\}$. Moreover, according to our assumption we have $c_p(\alpha^*_{p+1}) < c_{p+1}(\alpha^*_{p+1})$. So we have

$$\begin{aligned}\phi_p(p, \mu^*, t) &= \alpha^*_p MP(\alpha^*) - c_p\left(\alpha^*_p\right) + t_p \\ &= \alpha^*_{p+1} MP(\alpha^*) - c_p\left(\alpha^*_{p+1}\right) + t_{p+1} \\ &> \alpha^*_{p+1} MP(\alpha^*) - c_{p+1}\left(\alpha^*_{p+1}\right) + t_{p+1} = \phi_{p+1}(p+1, \mu^*, t)\end{aligned}$$

This establishes part 1. For part 2, again from (23), we have $\alpha^*_p MP(\alpha^*) - c_p\left(\alpha^*_p\right) + t_p = \alpha^*_{p+1} MP(\alpha^*) - c_p\left(\alpha^*_{p+1}\right) + t_{p+1}$. Rearrangement gives us $t_{p+1} = t_p + c_{p+1}\left(\alpha^*_{p+1}\right) - c_p\left(\alpha^*_p\right) - MP(\alpha^*)[\alpha^*_{p+1} - \alpha^*_p]$. From (11) and Proposition 2.4, we know that $MP(\alpha^*) = c'_p(\alpha^*_p)$. Thus we can write $t_{p+1} = t_p + c_{p+1}\left(\alpha^*_{p+1}\right) - c_p\left(\alpha^*_p\right) - c'_p(\alpha^*_p)[\alpha^*_{p+1} - \alpha^*_p]$. The strict convexity of $c(\cdot)$ implies $c_p\left(\alpha^*_{p+1}\right) - c_p\left(\alpha^*_p\right) - c'_p(\alpha^*_p)[\alpha^*_{p+1} - \alpha^*_p] > 0$. But by Observation 2.2, $c_{p+1}\left(\alpha^*_{p+1}\right) > c_p\left(\alpha^*_{p+1}\right)$.

Hence, $c_{p+1}\left(\alpha^*_{p+1}\right) - c_p\left(\alpha^*_p\right) - c'_p(\alpha^*_p)[\alpha^*_{p+1} - \alpha^*_p] > 0$, which gives us the desired result that $t_{p+1} > t_p$ for any $p = 1, 2, \ldots, n-1$. ■

Proof of Lemma 3.3: Recall from (18) that if the transfer vector is t, then $\phi_p(p, \mu^*, t) = \alpha^*_p MP(\alpha^*) - c_p\left(\alpha^*_p\right) + t_p$ and $\phi_p(q, \mu^*, t) = \alpha^*_q MP(\alpha^*) - c_p\left(\alpha^*_q\right) + t_q$. The fact that α^*_p is the unique best response to μ^* for a type p agent in the game $\hat{F}$ defined by (8) implies $\alpha^*_p MP(\alpha^*) - c_p\left(\alpha^*_p\right) > \alpha^*_q MP(\alpha^*) - c_p\left(\alpha^*_q\right)$.

Moreover, as $p > q$, $t_p > t_q$ by Lemma 3.2(2). Combining these arguments, we obtain $\alpha^*_p MP(\alpha^*) - c_p\left(\alpha^*_p\right) + t_p > \alpha^*_q MP(\alpha^*) - c_p\left(\alpha^*_q\right) + t_q$, or $\phi_q(q, \mu^*, t) > \phi_q(p, \mu^*, t)$.

Thus, (23) ensures that agents of type p do not have the incentive to claim to be of types $p+1$, $p+2$ etc (Lemma 3.1). That same condition also implies $t_1 < t_2 < \cdots < t_n$, i.e. Lemma 3.2(2), which then implies the present result that such agents will also not claim to be $q < p$. Therefore, if (23) is satisfied, no agent has any incentive to misrepresent type. ■

Proof of Proposition 3.4: Consider an arbitrary transfer scheme $\hat{t} \neq t^*$ that satisfies the IC constraints (19) and the budget balance condition. By (18), the payoff of an agent of type p who claims to be of type q under the transfer scheme $\hat{t}$ is $\phi_p(q, \mu^*, \hat{t}) = \alpha^*_q MP(\alpha^*) - c_p\left(\alpha^*_q\right) + \hat{t}_q$. It is easy to see that only difference between payoffs under $\hat{t}$ and t^* is the transfers, i.e. $\phi_p(p, \mu^*, \hat{t}) > (=) < \phi_p(p, \mu^*, t^*)$ if and only if $\hat{t}_p > (=) < t^*_p$.

First we make a few observations about the relation between t^* and $\hat{t}$. Since $\hat{t}$ is incentive compatible, it also satisfies the inequality (19). Thus, $\phi_p(p, \mu^*, \hat{t}) \geq \phi_p(p+1, \mu^*, \hat{t})$ or $\alpha^*_p MP(\alpha^*) - c_p\left(\alpha^*_p\right) + \hat{t}_p \geq \alpha^*_{p+1} MP(\alpha^*) - c_p\left(\alpha^*_{p+1}\right) + \hat{t}_{p+1}$ holds for all $p \in \{1, 2, \cdots, n-1\}$. To simplify notation, we denote $\lambda_{s,t} = \left[\alpha^*_s MP(\alpha^*) - c_s\left(\alpha^*_s\right)\right] - \left[\alpha^*_t MP(\alpha^*) - c_s\left(\alpha^*_t\right)\right]$. The notation $\lambda_{s,t}$ represents the difference in the payoff of a type s agent at the efficient state when he announces his type truthfully and the payoff that type s agent would receive when he announces some other type t. Now using this notation we can then rewrite the preceding inequality $\phi_p(p, \mu^*, \hat{t}) \geq \phi_p(p+1, \mu^*, \hat{t})$ as $\hat{t}_{p+1} \leq \hat{t}_p + \lambda_{p,p+1}$.[17] We now divide our proof into smaller claims.

[17] The t^* transfer vector satisfies this relation with equality, i.e. $t^*_{p+1} = t^*_p + \lambda_{p,p+1}$ for all $p \in \{1, 2, \ldots, n-1\}$.

(i) **Claim 1**: Suppose $\hat{t}_k > t_k^*$ for some $k \in \mathcal{P}$ then $\hat{t}_p > t_p^*$ for all $p \in \{1, 2, \cdots, k-1\}$.

Proof: From the above arguments we know that $\hat{t}_k \leq \hat{t}_{k-1} + \lambda_{k-1,k}$. Let $\hat{t}_k = t_k^* + \epsilon_k$ where $\epsilon_k > 0$. This implies, $t_k^* + \epsilon_k \leq \hat{t}_{k-1} + \lambda_{k-1,k}$. Now substitute the value of t_k^* in terms of t_{k-1}^*. We get $t_{k-1}^* + \lambda_{k-1,k} + \epsilon_k \leq \hat{t}_{k-1} + \lambda_{k-1,k} \Rightarrow t_{k-1}^* < \hat{t}_{k-1}$. We can apply this argument inductively to obtain the desired result.

(ii) **Claim 2**: Suppose $\hat{t}_k < t_k^*$ for some $k \in \mathcal{P}$ then $\hat{t}_p < t_p^*$ for all $p \in \{k+1, k+2, \ldots, n\}$.

Proof: We know that $\hat{t}_{k+1} \leq \hat{t}_k + \lambda_{k,k+1}$. Let $\hat{t}_k = t_k^* - \delta_k$ where $\delta_k > 0$. This implies, $\hat{t}_{k+1} \leq t_k^* - \delta_k + \lambda_{k,k+1}$. Now substitute the value of t_k^* in terms of t_{k+1}^*. We get $\hat{t}_{k+1} \leq t_{k+1}^* - \lambda_{k,k+1} - \delta_k + \lambda_{k,k+1} \Rightarrow \hat{t}_{k+1} < t_{k+1}^*$. We can apply this argument inductively to obtain the desired result.

(iii) **Claim 3**: If $\hat{t} \neq t^*$ then agents can be partitioned into at most three sets L, M and R where $L = \{1, 2, \ldots, l\}$, $M = \{l+1, l+2, \ldots, r-1\}$[18] and $R = \{r, r+1, \ldots, n\}$ for some $1 \leq l < r \leq n$ such that $\hat{t}_p > t_p^*$ for all $p \in L$, $\hat{t}_p = t_p^*$ for all $p \in M$ and $\hat{t}_p < t_p^*$ for all $p \in R$. The set L and R are always non-empty.

Proof: Define $\tilde{l} = \text{Max}\{p \in \mathcal{P} \mid \hat{t}_p > t_p^*\}$. By definition of $\tilde{l}$, there exists no $p > \tilde{l}$ such that $\hat{t}_p > t_p^*$. Moreover, according to Claim 1, for all $p = 1, 2, \ldots, \tilde{l} - 1$ we have $\hat{t}_p > t_p^*$. Hence, $l = \tilde{l}$ and the set $L = \{1, 2, \ldots, \tilde{l}\}$. Similarly, define $\tilde{r} = \text{Min}\{p \in \mathcal{P} \mid \hat{t}_p < t_p^*\}$. By definition of $\tilde{r}$ there exists no $p < \tilde{r}$ such that $\hat{t}_p < t_p^*$. Moreover, according to Claim 2, for all $p = \tilde{r}+1, \tilde{r}+2, \ldots, n$ we have $\hat{t}_p < t_p^*$. Hence, $r = \tilde{r}$ and the set $R = \{\tilde{r}, \tilde{r}+1, \ldots, n\}$. It is easy to see that $\tilde{l} < \tilde{R}$. Define set $M = \mathcal{P} \setminus L \cup R$. It is obvious that if $p \in M$ then $\hat{t}_p = t_p^*$ and $\tilde{l} < p < \tilde{r}$. Thus we obtain $M = \{\tilde{l}+1, \tilde{l}+2, \ldots, \tilde{r}-1\}$.

Now we argue that both sets L and R are non-empty. Without loss of generality, let $L = \emptyset$. This implies that $\hat{t}_p \leq t_p^*$ for all $p \in \mathcal{P}$ and strict inequity must hold for at least some $q \in \mathcal{P}$ otherwise $\hat{t} \equiv t^*$ which contradicts our hypothesis that $\hat{t} \neq t^*$. But this will imply that $\sum_{p=1}^{n} \hat{t}_p < \sum_{p=1}^{n} t_p^* = T(\mu^*)$ which contradicts our hypothesis that transfer $\hat{t}$ is budget balanced. Hence set L is non-empty. Virtually a similar argument can establish that if set $R = \emptyset$ then it implies $\sum_{p=1}^{n} \hat{t}_p > \sum_{p=1}^{n} t_p^* = T(\mu^*)$. This again contradicts our hypothesis. So, set R is also non-empty. This completes the proof.

(iv) **Claim 4**: Transfer t^* minimizes the variance.

Proof: Suppose not and there exists a budget balanced incentive compatible transfer $\hat{t}$ at the efficient state which has a lower variance than t^*. Claim 3 implies that the only possibility to decrease the variance is to increase payoffs of lower types at the cost of high cost agents. Define sets L, M and R of agents whose transfers or payoffs have increased, remain same and decreased in the new transfer $\hat{t}$ than t^*. Claim 3 again implies that $L = \{1, 2, \ldots, l\}$, $M = \{l+1, l+2, \ldots, r-1\}$ and $R = \{r, r+1, \ldots, n\}$ for some $1 \leq l < r \leq n$.

[18] Set M could be empty and in that case we have $r = l + 1$.

To simplify the notations we define $U_p = \phi_p(p, \mu^*, t^*)$ and $\hat{U}_p = \phi_p(p, \mu^*, \hat{t})$ for all $p \in \mathcal{P}$. According to our hypothesis we denote $\hat{U}_p - U_p = \epsilon_p > 0$ for all $p \in L$ and $U_p - \hat{U}_p = \delta_p > 0$ for all $p \in R$. Our earlier observations imply that $\hat{t}_p - t^*_p = \epsilon_p > 0$ for all $p \in L$ and $t^*_p - \hat{t}_p = \delta_p > 0$ for all $p \in R$. Using Claims 1 and 2, it can be easily shown that $\epsilon_l \leq \epsilon_{l-1} \ldots \leq \epsilon_1$ and $\delta_r \leq \delta_{r+1} \ldots \leq \delta_n$. Budget balance would require that $\sum\limits_{p=1}^{l} m_p\epsilon_p - \sum\limits_{p=r}^{n} m_p\delta_p = 0$. Now

$$
\begin{aligned}
V(\mu^*, \hat{t}) &= \sum_{p\in\mathcal{P}} m_p\big(\hat{U}_p\big)^2 - \big(\bar{F}(\mu^*)\big)^2 \\
&= \sum_{p\in L} m_p\big(U_p+\epsilon_p\big)^2 + \sum_{p\in C} m_p\big(U_p\big)^2 + \sum_{p\in R} m_p\big(U_p-\delta_p\big)^2 - \big(\bar{F}(\mu^*)\big)^2 \\
&= V(\mu^*, t^*) + \sum_{p\in L} m_p\epsilon_p^2 + \sum_{p\in R} m_p\delta_p^2 + 2\Big[\sum_{p\in L} m_pU_p\epsilon_p - \sum_{p\in R} m_pU_p\delta_p\Big] \\
&> V(\mu^*, t^*) + \sum_{p\in L} m_p\epsilon_p^2 + \sum_{p\in R} m_p\delta_p^2 + 2\Big[U_l\sum_{p\in L} m_p\epsilon_p - U_r\sum_{p\in R} m_p\delta_p\Big] \\
&> V(\mu^*, t^*)
\end{aligned}
$$

The first and second inequality follow from the facts $U_1 > U_2 > \ldots > U_n$ and $\sum\limits_{p=1}^{l} m_p\epsilon_p - \sum\limits_{p=r}^{n} m_p\delta_p = 0$. Thus, it implies that any arbitrary budget balanced and incentive compatible transfer scheme must produce a higher variance than the transfer scheme t^*.

The order $t^*_1 < t^*_2 < \cdots < t^*_n$ follows from Lemma 3.2(2). The order $\phi_1(1, \mu^*, t^*) > \phi_2(2, \mu^*, t^*) > \cdots > \phi_n(n, \mu^*, t^*)$ follows from Lemma 3.2(1). Hence, individual rationality will be satisfied if $\phi_n(n, \mu^*, t^*) > 0$. To see why this holds, we now establish another claim.

(v) **Claim 5**: The transfer vector t^* maximizes the lowest post redistribution payoff, i.e. the payoff of type n agents.

Proof: By part 1 of Lemma 3.2, type n has the lowest post redistribution payoff under any incentive compatible transfer vector t. Now consider vectors t^* and $\hat{t}$, both satisfying incentive compatibility and budget balance. Suppose the claim is not true and transfer vector $\hat{t}$ can do better. This is possible only if $\hat{t}_n > t^*_n$. But Claim 1 then implies that $\hat{t}_p > t^*_p$ for all $p \geq 1$. This implies that $\sum\limits_{p=1}^{n} \hat{t}_p > \sum\limits_{p=1}^{n} t^*_p = T(\mu^*)$, which means transfer $\hat{t}$ is not budget balanced. We, therefore, arrive at a contradiction. Hence, the claim is true.

Note that the equal redistribution transfer scheme $t_p = T(\mu^*)$, for all $p \in \{1, 2, \cdots, n\}$ also satisfies incentive compatibility and budget balance. Hence, by Claim 5, $\phi_n(n, \mu^*, t^*) > \phi_n(n, \mu^*, T(\mu^*))$. But by (16), $\phi_n(n, \mu^*, T(\mu^*)) = \alpha^*_p MP\left(\alpha^*\right) - c_p\left(\alpha^*_p\right) + T(\mu^*) > 0$. Hence, individual rationality is satisfied. ■

References

[1] Arrow K (1979) The Property Rights Doctrine and Demand Revelation Under Incomplete Information. in Economics and Human Welfare, ed. by M. Boskin. New York: Academic Press.

[2] Chakrabarti A, Lahkar R (2018) An Evolutionary Analysis of Growth and Fluctuations with Negative Externalities. Dyn. Games Appl. 2018:733–760.

[3] Clarke E (1971) Multi-part pricing of public goods. Public Choice 11:17–23.

[4] Corchón L (1994) Comparative statics for aggregative games the strong concavity case. Math. Soc. Sci. 28:151–165.

[5] d'Aspremont C, Gérard-Varet L–A (1979) Incentives and incomplete information. J. Public Econ. 11: 25-45.

[6] Elliott M, Golub B, Jackson MO (2014) Financial networks and contagion. Am Econ Rev 104:3115—53

[7] Green J, Laffont J-J (1979) Incentives in public decision-making. Elsevier North-Holland

[8] Groves T (1973) Incentives in teams. Econometrica 41:617–631.

[9] Lahkar R (2019) Elimination of non-individualistic preferences in large population aggregative games. J. Math. Econ. 84:150–165.

[10] Lahkar R, Mukherjee S (2019) Evolutionary Implementation in a Public Goods Game. J. Econ. Theory 181:423–460.

[11] Lahkar R, Mukherjee S (2020) Dominant Strategy Implementation in a Large Population Public Goods Game. Econ. Lett. 197:109616.

[12] Lahkar R, Mukherjee S (2021) Evolutionary Implementation in Aggregative Games. Math. Soc. Sci. 109:137–151.

[13] Ohseto S (2006) Characterizations of Strategy-Proof and Fair Mechanisms for Allocating Indivisible Goods. Econ. Theory 29:111–121.

[14] Pápai S (2003) Groves Sealed Bid Auctions of Heterogeneous Objects with Fair Prices. Soc. Choice Welf. 20:371–385.

[15] Rawls J (1999) A theory of justice: Revised edition. Harvard university press.

[16] Sandholm WH (2001) Potential games with continuous player sets. J. Econ. Theory 97:81–108.

[17] Sandholm WH (2002) Evolutionary Implementation and Congestion Pricing. Rev. Econ. Stud 69:667–689.

[18] Sandholm WH (2005) Negative Externalities and Evolutionary Implementation. Rev. Econ. Stud 72:885–915.

[19] Sandholm WH (2007) Pigouvian Pricing and Stochastic Evolutionary Implementation. J. Econ. Theory 132:367–382.

[20] Smith MJ (1984) The Stability of a Dynamic Model of Traffic Assignment—An application of a Method of Lyapunov. Transp. Sci. 18:245-–252.

[21] Sokoloff KL, Engerman SL (2000) Institutions, Factor Endowments, and Paths of Development in the New World. J. Econ. Perspect. 14:217–32.

[22] Sprumont Y (2013) Constrained-optimal Strategy-proof Assignment: Beyond the Groves mechanisms. J. Econ. Theory 148:1102–1121.

[23] Tadenuma K, Thomson W (1995). Games of Fair Division. Games Econ. Behav. 9:191–204.

[24] Vickrey W (1961) Counterspeculation, auctions, and competitive sealed tenders. J. Finance 16:8–37.