JOURNAL OF ECONOMIC THEORY **71**, 324–348 (1996)
ARTICLE NO. 0124

Egalitarianism and Incentives*

Debraj Ray

Boston University, Boston, Massachusetts 02215 and
Instituto de Análisis Económico (CSIC), 08193 Bellaterra, Barcelona

and

Kaoru Ueda

Boston University, Boston, Massachusetts 02215

Received May 2, 1995; revised March 14, 1996

A group of agents is collectively engaged in a joint productive activity. Each agent supplies an observable input, and output is then collectively shared among the members. A Bergson–Samuelson welfare function defined on individual utilities describes the social values of the agents. However, individual actions are taken on a selfish basis. The collective decision cannot be precommitted, and is made (after inputs are chosen) to maximize welfare conditional on the input decisions. This leads to inefficiency. The aim of this paper is to show formally that, contrary to popular belief, the degree of inefficiency *decreases* in the extent of egalitarianism embodied in the social welfare function. *Journal of Economic Literature* Classification Numbers: C72, D30, D63, P13.

1. INTRODUCTION

A social planner's concern for egalitarianism might lead to a dilution of incentives, and therefore a loss in efficiency. This intuitive notion has been explored in a number of different contexts.[1]

* Support under Grant PB90-0172 from the Ministerio de Educación y Ciencia, Government of Spain, and Grant SBR-9414114 from the National Science Foundation is gratefully acknowledged. We are grateful to Abhijit Banerjee, Arindam Das Gupta, Glenn Loury, and Robert Rosenthal for useful comments on an earlier version. We thank an anonymous referee for detailed comments on a first draft.

[1] The literature on the subject is broad-ranging and varied enough to preclude all but a cursory treatment in this short paper. For general discussions of the equity-efficiency tradeoff, see, e.g., Okun [14]. In the game-theoretic context, see, e.g., Dutta and Ray [15] and Geanakoplos and Nalebuff [7]. In the context of risk-bearing, see, e.g., Loury [10]. In the context of resource allocation in poor economies, see, e.g., Mirrlees [12] and Dasgupta and Ray [4].

0022-0531/96 $18.00

In an important class of situations, the efficiency loss arises because the planner cannot credibly commit to a future course of action, such as the decision *not* to tax an individual or group making efficiency-enhancing investments. So these investments are not made, or more generally, are undersupplied. For a discussion of similar problems, see Bergstrom [2], Lindbeck and Weibull [9], and (in a slightly different context) Bernheim and Stark [3].[2]

The issue that we address here is somewhat different. Specifically, in a setting with many agents, we wish to examine the connection between the extent of *egalitarianism* embodied in the social welfare function, and the consequent *degree* of inefficiency that results. As stated, these phrases are vague. They will be made precise in the formal analysis.

Such an analysis would possibly be entirely devoid of interest were it not for the fact that the results contrast sharply with the standard argument discussed above. Specifically, while a concern for egalitarianism (coupled with an inability to precommit), does generate an efficiency loss, we will argue that this loss becomes *smaller* as the degree of egalitarianism increases. In the extreme case of Rawlsian social preferences, the efficiency loss vanishes altogether. The goal of this paper is to make these points in a precise way.

Consider, then, the following class of situations. A group of agents is collectively engaged in a joint production activity, where the output from production is to be distributed among the members of the group. The agents (represented by a social planner, perhaps) are interested in maximizing the value of a Bergson–Samuelson social welfare function defined on their own utilities. However, while this welfare function represents their *social values*, individual actions are taken on an entirely selfish basis. Suppose that to achieve the desired outcome, each agent must take an observable action, followed by some *collective* action—the "social planner's move". Suppose, moreover, that a collective action (contingent on individual decisions) *cannot* be credibly committed in advance.[3]

Specifically, define a *soft mechanism* to be one that specifies a second best division of the output (relative to the social welfare function) conditional

[2] These and other studies draw their inspiration from the "Rotten Kid Proposition" of Becker [1]. The problem in the various examples studied in these papers is to examine the conditions under which altruism, coupled with the inability to precommit actions, might give rise to inefficient behavior.

[3] A central example is the problem of allocating funds from a central government to different state governments. Typically, each state carries out a number of different expenditure programs, which are financed from central funds and state revenues. The center would like to channel proportionately greater funds to the poorer states. At the same time, the center would like to induce each state to carry out activities that will raise *per capita* income in that state. But the former goal places limits on the punishments that a center can credibly impose on a state for not taking actions to further the latter objective.

on *every* possible input vector. We wish to compare the resulting equilibria of the "soft game" so induced, with the first best under the very same welfare function.

Proposition 1 states that soft mechanisms always lead to *underproduction* relative to the first best. The reader will not find this surprising, though in this context the result has not been proved before, and indeed, is not generally true.[4] But overall, a familiar intuition is confirmed.

The next two results make the main points of the paper. We consider, first, the case of extreme egalitarianism, as embodied in a Rawlsian social welfare function. It turns out (Proposition 2) that in this case, there is a unique equilibrium of the soft game, *and the outcome coincides exactly with the Rawlsian first best*. There is no inefficiency. This striking result cannot be rationalized by the argument that the Rawlsian first best is very different from first best outcomes under other welfare functions. In a large class of symmetric situations, the first best is independent of the welfare function, as long as the latter is symmetric.

Translating the extreme Rawlsian result into a comparative statics property with respect to changing egalitarianism is our next task. We first provide a natural partial ordering on welfare functions that correspond to an increased preference for egalitarianism. We show that well-known parameterizations of increased inequality-aversion, such as the Atkinson class of welfare functions, satisfy this ordering. Our general class admits additive utilitarianism as one extreme point of the ordering, and Rawlsian social preferences at another. Of course, to each element in this class is also associated a soft mechanism. We show that as egalitarianism increases, the degree of inefficiency under the corresponding soft mechanisms is monotonic in the following sense: for each (symmetric) equilibrium of the soft game under a given degree of egalitarianism, and for each preassigned greater degree of egalitarianism, there is an equilibrium in the latter soft game with a lower degree of underproduction. This is Proposition 3.

Proofs of all the propositions are in Section 6.

A number of features underlying the soft mechanism require, perhaps, additional discussion. Chief among these is the assumption that past sunk effort levels are not ignored, but taken into account in subsequent welfare calculations. Some discussion of this (and other issues) will be found in Section 4, but some summary remarks may be useful here.

The point of the paper is that a concern for egalitarian outcomes enhances incentives, *provided that the actions that are subject to moral hazard are taken into account in the evaluation of social welfare ex post*. It is often the case that such past actions escape the memory of *ex post* evaluations, or

[4] For example, Proposition 1 fails to hold if individual utilities are linear in consumption. See below.

worse still, are imperfectly observable. In that case, egalitarianism hinders the taking of such actions by not rewarding them appropriately *ex post*. We argue, then, that it is precisely the issue of whether past actions are considered in *ex post* evaluations that should be central to the study of egalitarianism and incentives.[5]

2. MODEL

2.1. *Technology and Individual Preferences*

Consider a group of n individuals ($n \geqslant 2$) producing a single output. Output is produced by the joint efforts of these individuals according to the production function

$$Y = F(\mathbf{e}) \tag{1}$$

where $\mathbf{e} = (e_1, ..., e_n)$ is a (nonnegative) vector of *efforts*. We assume

(A.1) *F is continuous and concave with $F(\mathbf{0}) = 0$ and differentiable whenever $F > 0$, with $\partial F(\mathbf{e})/\partial e_i \equiv F_i(\mathbf{e}) > 0$.*

Each individual has preferences over pairs of consumption (c) and leisure (l). We assume

(A.2) *For each i, preferences are representable by an increasing utility function $u^i\colon \mathbb{R}^2_+ \to \mathbb{R}$ which is $\mathscr{C}^2$ and strictly quasiconcave, with $(u^i_c(c, l), u^i_l(c, l)) \geqslant 0$ whenever $u(c, l) > u(0, 0)$.[6] Moreover, leisure is a normal good.[7]*

Each individual has a labour endowment L_i. Thus, for each i,

$$L_i = l_i + e_i \tag{2}$$

Let $\mathscr{E}$ be the product of intervals $[0, L_i]$ over all i. Note that whenever $e \in \mathscr{E}$, we shall have l_i given by (2), for $i = 1, ..., n$. Keeping this in mind, define an *outcome* as a pair $(\mathbf{c}, \mathbf{e}) = ((c_1, ..., c_n); (e_1, ..., e_n))$, such that $c \geqslant 0$, $e \in \mathscr{E}$, and

$$c_1 + c_2 + \cdots + c_n = F(\mathbf{e}). \tag{3}$$

[5] There is also the question of how egalitarianism affects incentives when there is incomplete information regarding the innate abilities of the agents involved. This question is beyond the scope of the present exercise, but is addressed in our forthcoming work.

[6] Interpret these as right-hand derivatives when c or l equals 0. A technical remark: it is well known that strict quasiconcavity, coupled with $\mathscr{C}^2$, does not guarantee that the determinant of each principal minor of the relevant bordered Hessian is non-zero. We use the definition in the stronger sense to rule out these degenerate cases. On these matters, see, e.g., Sydsaeter [15].

[7] The assumption that leisure is a normal good implies the restriction $u_{ic} u^i_{cl} - u^i_l u^i_{cc} > 0$.

2.2. *Social Preferences*

The social planner is presumed to possess Bergson–Samuelson preferences satisfying standard restrictions:

(A.3) (i) *Society's preferences are representable by a* $\mathscr{C}^2$ *welfare indicator* W: $\mathbb{R}_+^{2n} \to \mathbb{R}$, *defined on 2n-tuples of consumption-leisure vectors. This welfare function is strictly quasiconcave. Moreover, for each* $\mathbf{l} \in \mathscr{E}$ *and each i,* c_i *is not an inferior good under the function* $W(\cdot, \mathbf{l})$.

(ii) *There exists a function* V (*which will be* $\mathscr{C}^2$ *by part* (i) *and* (A.2)) *such that* $W(c, l) = V(u^1(c_1, l_1), ..., u^n(c_n, l_n))$. *Moreover,* $\partial V(u)/\partial u^i \equiv V_i(u) > 0$ *for all i.*

2.3. *First Best*

The planner would like to maximize social welfare. This is achieved by an outcome $(\mathbf{c}, \mathbf{e})$ that maximizes $W(\mathbf{c}, \mathbf{L}\text{-}\mathbf{e})$, where $\mathbf{L} \equiv (L_1, L_2, ..., L_n)$ is the vector of labor endowments. From Assumptions 1 and 3 it follows that there is a *unique* outcome $(\mathbf{c}^*, \mathbf{e}^*)$ that solves this problem. Call this outcome the *first best.*

Some minor restrictions on the first best are summarized in the following assumption.[8]

(FB) *There is at least one individual i such that* $e_i^* > 0$ *and* $l_i^* \equiv L_i - e_i^* > 0$. *Moreover, the first best output is larger than the optimal output that would be chosen by any one individual acting on his own*: $F(\mathbf{e}^*) > F(0, ..., 0, \hat{e}_j, 0, ..., 0)$, *where* $\hat{e}_j$ *solves* $\max_{0 \leqslant e_j \leqslant L_j} u^j(F(0, ..., 0, e_j, 0, ..., 0), L_j - e_j)$. *Finally, under the first best, each individual obtains at least as much utility as from total inaction* (*which yields* $u^i(0, L_i)$).

2.4. *The Soft Mechanism*

The idea of a no-commitment mechanism, or *soft mechanism* is based on the postulate that the planner cannot avoid maximizing welfare *ex post*, even though this may be detrimental to the maximization of welfare *ex ante*. Formally, for each *given* vector $\mathbf{e} \in \mathscr{E}$, consider the problem

$$\max_{\mathbf{c}} W(\mathbf{c}, \mathbf{L} - \mathbf{e}) \tag{4}$$

subject to the constraint that $(\mathbf{c}, \mathbf{e})$ must be an outcome.

By (A.1) and (A.3), there is a unique vector $c(\mathbf{e})$ that solves (4). We will refer to $(c(\mathbf{e}), \mathbf{e})$ as an *ex post outcome.* This captures, in an extreme way, the inability to impose arbitrary punishments on deviants. The lack of

[8] Not all the restrictions are used in all the propositions, but the conditions are so minor that we do not feel it useful to separate this list into separate components.

commitment creates limitations on the incentive mechanisms that can be used to stimulate production.

The collection of ex post outcomes $(c(\mathbf{e}), \mathbf{e})$, for $\mathbf{e} \in \mathscr{E}$, induces a *soft game* in the obvious way: Player i chooses $e_i \in [0, L_i]$. If the vector $\mathbf{e}$ is chosen, i consumes $c_i(\mathbf{e})$, thus generating the payoff $u^i(c_i(\mathbf{e}), L_i - e_i)$.

2.5. *An Additional Restriction*

We now have sufficient terminology to introduce an additional joint assumption on technology and preferences, which will be used to derive one of the main results. To motivate this assumption first consider a standard property of the first best outcome, which can be easily verified: If $e_i^* < L_i$,

$$u_c^i(c_i^*, l_i^*)\, F_i(e^*) \leqslant u_l^i(c_i^*, l_i^*) \tag{5}$$

(indeed, with equality holding if $e_i^* > 0$).

The intuition is simple. If this inequality did not hold, i's effort could be raised a little with all the additional output being credited to him. His utility would be higher, with every other utility remaining constant. Social welfare goes up, a contradiction to the fact that we have a first best outcome to start with.

The additional assumption that we wish to make is related closely to (5). Specifically, we suppose:

(A.4) *For each ex post outcome* $(c(\mathbf{e}), \mathbf{e})$ *such that* $F(\mathbf{e}) \geqslant F(\mathbf{e}^*)$, *there is i with* $e_i > 0$, *and*

$$u_c^i(c_i, l_i)\, F_i(\mathbf{e}) \leqslant u_l^i(c_i, l_i) \tag{6}$$

where $c_i(\mathbf{e}) \equiv c_i > 0$.

This assumption looks plausible, because as output moves above the first best, we would expect that marginal products do not increase, while the marginal rate of substitution between consumption and leisure tilts in favor of leisure. Thus if (5) already holds at the first best, we expect this relationship to be maintained (for at least one individual) for outcomes with higher output.

However, it is only fair to point out (A.4) is *not* automatically implied by (A.1)–(A.3). However, experimentation with different functional forms suggests that it is implied by a large subclass of welfare functions, individual preferences and production technologies.

For instance, suppose (in addition to (A.1)–(A.3)) that individual utilities are separable as the sum of concave functions of consumption and leisure, that total output is some concave, smooth function of the *sum* of individual efforts, and that the social welfare function has a separable and concave representation in utilities.

Consider some ex post outcome $(\mathbf{c}(\mathbf{e}), \mathbf{e})$, distinct from the first best, but with the property that at least as much output is being produced as in the first best: $F(\mathbf{e}) \geqslant F(\mathbf{e}^*)$. There must be *some* individual i with $e_i > e_i^*$. In the case under consideration, it will be the case that at least *one* such individual gets $c_i(\mathbf{e}) \geqslant c_i^*$ (remember that at least as much output is being produced). Furthermore, since output is no lower, marginal product cannot have increased relative to the first best. Putting all this together with (5) and using the convexity properties of preferences and technology, it should be the case that

$$u_c^i(c_i(\mathbf{e}), L_i - e_i)\, F_i(\mathbf{e}) \leqslant u_l^i(c_i(\mathbf{e}), L_i - e_i)$$

for this individual, which is (6).

Consider, therefore, the following formalization of the separable case:

(A.4*) (i) *u^i is additively separable in* (c, l), (ii) $F(\mathbf{e})$ *is of the form* $f(e_1 + \cdots + e_n)$, for *some* $f: \mathbb{R}_+ \to \mathbb{R}_+$, and (iii) V is separable in individual utilities.

Our motivating argument above shows that under (A.1)–(A.3), (A.4*) implies (A.4).

3. EGALITARIANISM AND INCENTIVES

3.1. *Egalitarianism Yields Underproduction...*

PROPOSITION 1. *Under* (A.1)–(A.3) *and* (FB), *the first best cannot be achieved as some equilibrium of the soft game. Moreover, if* (A.4) *holds, every equilibrium of the soft game must involve underproduction.*

Proposition 1 verifies an intuitive conjecture: when a team cannot precommit to adequately punish deviants, the first best cannot be achieved. Specifically, in this model there is underproduction relative to the first best.

It is perhaps worth mentioning that while Proposition 1 appears intuitive (especially because it is related to other inefficiency notions, such as the Marshallian inefficiency of sharecropping, or the holdup problem), it is far from being obviously true in the present context. For instance, if individual utility has a separable linear representation in consumption, the proposition is, in general, false: there are equilibria that attain the first best.[9] Moreover, without (A.4), whether there is underproduction or not is an open question. Most significantly, the assumptions of Proposition 1 do not cover the case of Rawlsian social preferences, and as we shall see in the next section, our proposition fails in this case.

It might help, therefore, to sketch the proof of Proposition 1 for a special case. Assume that individual preferences have a strictly concave representation that is separable in consumption and leisure, and that given this representation, the Bergson–Samuelson welfare function has a representation that is strictly concave and separable in individual utilities. Furthermore, assume that total output is a concave, increasing function f of the *sum* of individual efforts. This is the case covered by (A.4*).

Now consider an ex post outcome $(\mathbf{c}, \mathbf{e})$ with *overproduction* relative to the first best. *Someone* must be putting in more effort relative to the first best. Call that someone i. Evaluated at the first best consumption vector, i must now have a higher marginal weight in the social welfare function than he had before. In the ex post outcome under consideration, therefore, i must enjoy more consumption, because total output is not lower. Putting all this information together with the convexity of preferences and technology, it is easy to see that

$$u_c^i f'(e_1 + \cdots + e_n) \leqslant u_l^i \tag{7}$$

which just means that (A.4) is automatically satisfied in this special case.

Now, suppose that i reduces his effort level a tiny bit. Perform the thought experiment of cutting i's consumption by *exactly* the resulting fall in output. By the inequality (7), this has a *nonnegative* first-order effect on i's utility. If the outcome is to be a Nash equilibrium, therefore, i's true consumption decline must be *at least* of this order. But now note that this decline must have a first-order effect on i's marginal utility of consumption, which rises. Moreover, by the supposition that i is no better off, the marginal welfare contribution of his utility is no lower. It is now easy to check that in this new situation, the first-order conditions for ex post welfare maximization are destroyed. For because i is being made to take at least the entire output loss (as we have argued above), there must be some other individual whose consumption is no lower. For this other person, exactly the reverse changes must occur in the marginal conditions. So, if the first order conditions were holding earlier (as they must have), they cannot be holding now! This contradicts ex post welfare maximization, and proves that the original outcome could not have been an equilibrium.

[9] Such utility functions are ruled out by the assumption that leisure is a normal good (see (A.2)). The reason such functions may permit the first best to be attained is discussed briefly after the statement of Proposition 2.

3.2. ... *But Not with a Rawlsian Planner*

Our main theme relates the degree of egalitarianism to the degree of efficiency failure. It will be convenient to begin by exploring the most extreme form of egalitarianism: Rawlsian social preferences. In a later section, we will carry out the more complicated exercise of varying the welfare function over different degrees of egalitarianism.

Observe that the Rawlsian case is *not* covered by Proposition 1. That proposition rests on a postulate that is seemingly so innocuous that we have not emphasized it in the discussion (though, of course, it is formally stated in (A.3)). It is that social welfare is *strictly* increasing in *every* utility level, for each vector of utilities. Rawlsian social preferences do not satisfy this condition. Indeed, this observation has striking consequences, as we shall see.

We first define Rawlsian preferences. To do so we need a benchmark comparison. This is summarized in the following assumption:

(R) *Society is indifferent between the complete inaction of any two individuals. Thus if Rawlsian welfare is written as*

$$V(\mathbf{u}) = \min_i u^i \tag{8}$$

utilities have already been normalized so that $u^i(0, L_i) = u^j(0, L_j)$ *for all* i *and* j.

The first best Rawlsian optimum involves the maximization of the expression in (8), subject to the constraints that $u^i = u^i(c_i, l_i)$ for all i and that $(\mathbf{c}, \mathbf{e})$ is an outcome. Just as before, there is a unique first best outcome; call it (c^*, e^*).

Ex post outcomes are defined exactly as they were earlier. It is easy to see that for each $e \in \mathscr{E}$, there is a unique consumption vector $c(\mathbf{e})$ that solves the ex post maximization problem. The collection-of ex post outcomes induces a soft game just as before, and we may study its equilibria.

PROPOSITION 2. *Under* (A.1), (A.2), (FB) *and* (R), *the Rawlsian first best is an equilibrium of the Rawlsian soft game. Furthermore, any equilibrium which gives at least one agent strictly more utility than the utility from inaction must be the Rawlsian first best outcome.*

Remarks. (1) One might wonder whether the strikingly positive result of Proposition 2 is due to the fact that the Rawlsian egalitarian optimum is so devoid of efficiency properties that we do not have any incentive problem at all to maintain it. But this is not the case. For instance, consider any special case of this model that is *symmetric* across all agents. It is easy to see that the first best outcome is invariant across all symmetric quasi-concave welfare functions, including the Rawlsian one.

(2) There is only one qualification in Proposition 2. To prove that an equilibrium must be the Rawlsian first-best, we assume that at least one individual receives strictly more utility than he receives from inaction. There may exist an equilibrium involving total inaction. This will happen if the technology has the property that $F(\mathbf{e}) = 0$ whenever $n-1$ components of $\mathbf{e}$ equal zero. On the other hand, if the technology is such that output depends on the sum of the efforts (and if right-hand marginal utilities are defined and strictly positive everywhere) then our qualification can be dispensed with.[10]

Thus, in contrast to standard intuition, extreme egalitarianism might actually have pleasing incentive properties. Egalitarianism applies not only to the choice of the social optimum, but in the treatment of deviants from the optimum. A greater concern for egalitarianism goes hand in hand with the ability to credibly mete out stronger punishments.

Apart from the technicalities, the proof of Proposition 2 is very simple and general. Under convexity of the feasible set and preferences, the Rawlsian criterion has the following property. All individual utilities move in the *same* direction from one ex post outcome to another. Consequently, the Rawlsian first best is always an equilibrium of the Rawlsian soft game. For if someone could improve his utility by a deviation, he would improve the utility of everyone else in the process. This would contradict the fact that the earlier outcome was first best.[11]

The second part of the result—that *every* equilibrium must be first best—is model-specific in two respects. The convexity and the differentiability features of the model must both be exploited. These are used to guarantee that if an allocation is *not* first best, then there is some small, *unilateral* change in someone's effort level that creates an ex post outcome with a higher Rawlsian value. Again, using the Rawlsian criterion and the equal-utility property yielded by convexity, the individual who makes the

[10] Our requirement that. the production function be differentiable at positive output is used in the proof. Concave, increasing production functions which produce no output when $n-1$ effort components equal zero are not (right hand) differentiable at 0, even though all partial (right hand) derivatives may exist at 0.

[11] It should be pointed out that this joint movement of utilities in the same direction is also a feature of the case in which utilities have a separable linear representation in consumption. To see this, let $u^i(c, l) = a_i c + b_i(l)$, where $a_i > 0$ and b_i is an increasing, smooth, concave function. Assume that $V(\mathbf{u}) = v_1(u_1) + \cdots + v_n(u_n)$, where each v_i is smooth, increasing, and strictly concave. Now look at the ex post outcomes of this model. If all consumptions are strictly positive, then the necessary and sufficient conditions characterizing ex post optimality are that for all i, j, $v_i'(u^i(c_i, l_i))\, a_i = v_j'(u^j(c_j, l_j))\, a_j$. The co-movement of utilities should now be apparent. So in this case as well, the first best can be supported as an equilibrium of the soft game.

change must participate in its benefits, thereby destroying the equilibrium possibilities of the given outcome.

3.3. *Changing Egalitarianism*

The result of the previous section, and the discussion following it, suggest an even stronger observation: that as the extent of egalitarianism increases, the degree of underproduction should monotonically decline. Proposition 2 would be the limiting case of such an observation. The purpose of this section is to demonstrate such a possibility.

We consider a *symmetric* version of our model, described formally as follows. First, any permutation of **e** produces the same output level as F(**e**). Second, every individual's preferences is represented by the same function u. Finally, the social welfare function is symmetric:

$$W(\mathbf{c}, \mathbf{l}) = V(u(c_1, l_1), ..., u(c_n, l_n)) \tag{9}$$

where V is symmetric. We will fix the individual representation u, and analyze the effect of changing egalitarianism by altering the form of V in a manner made precise below.

The following assumption (in addition to those already maintained) will be made on the fixed cardinal representation, u, of individual preferences.

(A.5) *There exists a cardinal representation such that u is strictly concave in c, and V is quasiconcave in* **u**.[12]

We begin by discussing how to compare the "degree of egalitarianism" among different V's, or more generally, among different social welfare orderings on vectors of individual utility. Our definition, while not formally requiring symmetric social welfare orderings, is best viewed in this background. Let S and S' be two social welfare orderings for utility vectors **u**. We will say that S' is *at least as egalitarian as* S if for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$, and for all i and j, the condition

$$u_k = v_k \quad \text{for} \quad k \neq i, j, \quad \mathbf{v} S \mathbf{u}, \quad \text{and} \quad |v_i - v_j| < |u_i - u_j| \tag{10}$$

implies $\mathbf{v} S' \mathbf{u}$.

That is, S' is more egalitarian than S if it prefers every redistribution plan to narrow the difference between two individuals' utility that S prefers as well. This can be viewed as an extension of the Pigou–Dalton principle for utility distributions (see, for example, Moulin [13]), applied to a comparison of any two social welfare orderings.

[12] The fix on a cardinal representation that is strictly concave is not necessary for the results (as long as (A.3) is maintained), but it greatly simplifies the writing of the proofs.

Some remarks on this last observation will clarify our definition. Observe that under our definition, a social welfare ordering S' is at least as egalitarian as the *utilitarian* welfare ordering if the condition

$$u_k = v_k \quad \text{for} \quad k \neq i, j, \quad u_i + u_j \leqslant v_i + v_j, \quad \text{and} \quad |v_i - v_j| < |u_i - u_j|$$

implies $\mathbf{v} S' \mathbf{u}$. Note that the condition above is just a specialization of (10), when S is the utilitarian ordering. Specializing even further so that $u_i + u_j = v_i + v_j$ in the above condition, we obtain the Pigou–Dalton principle for utility distributions. In other words, a social welfare ordering which is at least as egalitarian as the utilitarian ordering satisfies the Pigou–Dalton principle. Our suggested relation is therefore an extension of this idea.

The argument above suggests that as we progressively move "toward egalitarianism" in this ordering, we obtain social welfare functions that are "more willing" to trade off total utility for interpersonal equality.

This definition and the subsequent discussion lead to two general observations. First, in the class of social welfare functions considered in this paper, every such function is at least as egalitarian as the utilitarian function.[13] Second, the Rawlsian function is at least as egalitarian as any of these social welfare functions.[14] In short, the utilitarian and the Rawlsian social welfare orderings are two extremes of the social welfare functions we are considering, in the spectrum of the degree of egalitarianism defined here.

In our model, social welfare orderings are given by the functions satisfying (A.2) and (A.3), which are differentiable. Let V and V' be two such *symmetric* functions. Then we can show that V' is at least as egalitarian as V if and only if

$$\text{For all } \mathbf{u} \text{ and for all } i, j \in \{1, ..., n\}, \qquad u_i \geqslant u_j \Leftrightarrow \frac{V'_i}{V'_j} \leqslant \frac{V_i}{V_j}. \tag{11}$$

We omit a proof of this result (it is available on request).[15] But the interpretation should be quite natural. When one individual j has a lower utility than another individual i, the "more egalitarian" social welfare function assigns a higher (relative) marginal welfare to i. The social marginal rate of

[13] This is because increasing, strictly quasiconcave welfare functions will indeed rank $\mathbf{v}$ over $\mathbf{u}$ whenever the specialization of condition (10) discussed in the text happens to hold.

[14] This is easily seen as follows. Suppose that for some social welfare ordering S, and utility vectors $\mathbf{u}$ and $\mathbf{v}$ condition (10) holds but nevertheless, $\mathbf{u}$ is preferred by the Rawlsian ordering to $\mathbf{v}$. This means, in particular that $\min\{u_i, u_j\} > \min\{v_i, v_j\}$. But because $|u_i - u_j| > |v_i - v_j|$, this implies that (u_i, u_j) *vector-dominates* (v_i, v_j). But now we have a contradiction to the assumption that $\mathbf{v} S \mathbf{u}$.

[15] It is worth noting that the asserted equivalence is an analogue of a result in Moulin [13, Lemma 2.2], which characterizes a differentiable social welfare function (called a "collective utility function") that satisfies the Pigou–Dalton principle for utility distributions.

substitution between i and j is tilted in favor of j by the more egalitarian welfare function.

To obtain a clearer idea of the partial ordering proposed here, consider two classes of social welfare functions. The first is the Atkinson family, which is represented by the form

$$\left[\sum_{i=1}^{n} u_i^{\rho} \right]^{1/\rho}, \tag{12}$$

where $\rho \in (-\infty, 1]$ and $\rho \neq 0$, and by $\prod_{i=1}^{n} u_i$ if $\rho = 0$. It is easy to check that higher degrees of egalitarianism correspond to lower values of ρ. In particular, the case $\rho = 1$ yields the utilitarian, and $\rho = -\infty$, the Rawlsian social welfare ordering.

The second class is the "constant absolute risk-aversion" family, which is given in the form

$$-\frac{1}{\eta} \sum_{i=1}^{n} \{\exp(-\eta u_i) - 1\}, \tag{13}$$

where $\eta \in (0, \infty)$. A higher value of η is associated with a more egalitarian function, and the two extreme values of η (0 and ∞) again give the utilitarian and the Rawlsian social orderings respectively.

Now, let us return to the model. An outcome is *symmetric* if all individuals work equally hard, and consume equally. In the symmetric model, symmetric effort by individuals induces a symmetric outcome, in which $\mathbf{c}(e, ..., e) = (c, ..., c)$ and $c = (1/n) F(e, ..., e)$. The first best (c^*, e^*) is also a symmetric allocation, and the allocation is independent of the social welfare function.

We will restrict our discussion to the class of symmetric equilibria of the soft game.

In what follows, it will be convenient to consider a modified soft game where all but the first individual choose the same level of effort. Then the *ex post* outcome assigns the same level of consumption for those providing this effort. Denote the generated levels of consumption by

$$(c_1(e_1, e; V), c(e_1, e; V)) \tag{14}$$

where V is the social welfare function determining consumption, and e and c are the common values for all but the first individual. The first individual's best response to e is the set of solutions to the problem of maximizing his indirect utility function with respect to $e_1 \in [0, L]$. Denote this set by $B(e; V)$. $B(e; V)$ is a compact-valued and upper hemicontinuous correspondence of e. We will need to assume:

(A.6) *$u(c_1(e_1, e; V), L - e_1)$ is single-peaked with respect to e_1.*

Then, $B(e; V)$ is convex-valued and the existence of a symmetric equilibrium for each V is guaranteed.[16]

By the *degree of underproduction* in any outcome, we refer to the difference between the first-best production level and the level of aggregate production under that outcome. The following proposition establishes that while the degree of underproduction is positive, the degree of underproduction must *fall* with rising egalitarianism.

PROPOSITION 3. *Assume a symmetric model satisfying* (A.1)–(A.3), (A.5), (A.6), *and* (FB). *Then every symmetric equilibrium of the soft game involves underproduction. Now consider two welfare functions V and V', where V' at least as egalitarian as V. Then, for each symmetric equilibrium under V, there exists a symmetric equilibrium under V' such that the degree of underproduction is not more for the latter.*

This verifies that the Rawlsian case discussed in Proposition 2 is not an exception or some quirky failure of continuity. For symmetric equilibria of the symmetric model, increased egalitarianism never increases the degree of underproduction.[17]

4. EXTENSIONS

We briefly consider some extensions, as well as possible objections to the setting in which these results have been derived.

1. *Sinking Past Differences.* In computing ex post optima, why do teams take account of the sunk efforts already incurred by its members? It may be argued that bygones are bygones, and that the social welfare function should ignore this. This point of view may be identified with the assertion that individual utilities "should" be separable in consumption and leisure, and that the social welfare function "should" be utilitarian. At least, that is the only way to justify the assertion if one sticks to the Bergson–Samuelson setting. Such particular functional forms are already accommodated as a special case of the paper, and if one insists on such an interpretation, an entire class of welfare functions (including the Rawlsian function) are simply removed from consideration.

[16] It should be noted that our maintained assumptions *do not* automatically yield (A.6). However, all parametric forms that we have tried satisfy this assumption.

[17] It is possible to find even stronger versions of this proposition in special cases. For instance, if the best-response correspondence is single-valued, and social welfare functions are in the Atkinson class, we can show that increased egalitarianism *strictly* lowers the maximal degree of underproduction. The qualification "maximal" is due to the fact that equilibrium may not be unique.

On the other hand, one might take the sinking of past differences as a primitive, simply asserting that ex post output is shared equally (in a symmetric model) *irrespective* of the welfare function. We find it difficult to see what might justify such an assumption. Consider any resource allocation problem. As long as resources are not allocated to everybody at exactly the same point of time, this point of view leads to absurd allocations.

If one were to admit such a structure for the sake of argument, inefficiencies will *always* arise, of course. But even then, it can be shown that as long as *some* weight is given to the past, Proposition 3 will continue to hold. In particular, the Rawlsian welfare function will exhibit the lowest degree of underproduction.

2. *Emotions and Credible Punishments.* One might proceed in the exactly the opposite direction to (1). Deviants that do not adhere to the desired outcome might enjoy an entirely different (and reduced) weight in the ex post welfare function. Emotions such as anger or social disapproval might induce such changes, and in so doing, lend credibility to punishments (for similar ideas, see, e.g., Frank [6]). It is to be expected that the degree of inefficiency will be lowered in the presence of these emotions. But even so, as long as the first best is thereby not automatically achieved, more egalitarian welfare functions will possess better efficiency properties.

3. *The Observability of Effort.* The results of this paper rest critically on the assumption that efforts are observable. We suspect that the results would extend to noisy observability, though we have not checked this. Of course, if efforts are not observable *at all*, the results cease to have any relevance.

4. *Repeated Relationships.* In a dynamic situation, teams would recognize that a departure from the soft mechanism will serve them well in the longer run. It may be of interest to study such repeated relationships. But as a prelude to that study, it is surely important to analyze the "one-shot" relationship without precommitment, which is exactly what we do here.

5. DISCUSSION

While the connections between egalitarianism and inefficiency has long been a subject of debate, there have been surprisingly few attempts to model the exact nature of the tradeoff. To be sure, the sources of various tradeoffs are manifold in nature. This paper investigates one potential source, and shows that a commonly held intuition is not valid, at least in this case.

Specifically, this paper studies the idea that egalitarianism fails to uphold proper incentives because credible punishments are thereby destroyed. This statement really has two parts to it: one is the familiar "dilemma of the Samaritan" induced by the inability to precommit. Organizations that cannot precommit, *yet derive their sense of goal-fulfillment or welfare from the welfare of its members*, are particularly prone to these potential inefficiencies. Indeed, organizational structures where the source of utility for the "principal" is directly opposed to that for the "agent" will certainly do better (in terms of efficiency) compared to the structures considered here.[18] This is the intuition upheld by Proposition 1. An inability to precommit the reward function results in inefficiency. While the *particular* result proved is, to our knowledge, new, there is nothing particularly surprising or novel about the underlying theme.

But this is only half of the idea. The second part goes further. It states that *egalitarian* organizations faced with the inability to precommit are doubly cursed: they are inefficient on the additional count that they cripple the incentive system (already weakened by the inability to precommit) even further. In this paper, we argue that this assertion is wrong. Increased egalitarianism *restores* incentives that are damaged by the lack of commitment (Proposition 3). Indeed, in the extreme case of Rawlsian egalitarianism, the precommitment and no-precommitment yield exactly the same first-best outcomes.

The results in this paper are provocative on two counts. First, they might inspire greater interest in a challenging and crucially important area of research: the connections between the ethic of equality and the yardstick of aggregate performance (such as GNP growth). The second aspect is one that we have not emphasized in this paper, but of great interest, we believe. This is the connection with the theory of implementation. By far the dominant approach in this literature presumes that the planner can implement outcomes without regard to ex post credibility. But there are many situations where it is natural to constrain mechanisms off the equilibrium path

[18] Indeed, in the usual principal–agent setup the problem is trivial. If efforts are observable, the organization is best off using a forcing contract. The contract demands that each individual must supply (at least) his first best effort level. If this effort level is indeed supplied by *all* individuals, all individuals receive the first best output division. If, however, these effort levels are not forthcoming, the contract promises dire consequences for the individuals who have deviated. These threats indeed implement the first best (at least as one equilibrium), and there is no incentive problem worth serious analysis. Authors such as Holmstrom (Holmstrom [8]) have argued that the existence of a residual claimant (capitalist, manager) creates a credible threat to carry out these punishments. such as retention of produced output, in the event of breach.

by the provision that they should not be suboptimal relative to the actions that have been taken and observed, and the social welfare function of the planner. This may have some bearing on narrowing the class of allowable mechanisms in implementation contexts.[19]

6. PROOFS

LEMMA 1. *Consider some ex post outcome* $(c(\mathbf{e}), \mathbf{e})$ *such that* $F(\mathbf{e}) \geqslant F(\mathbf{e}^*)$, *with the additional property that for some individual i,*

$$u_c^i(c_i, l_i) F_i(\mathbf{e}) \leqslant u_l^i(c_i, l_i). \tag{15}$$

Then such an outcome cannot be an equilibrium of the soft game.

Proof. Suppose that $(c(\mathbf{e}), \mathbf{e}) \equiv (\mathbf{c}, \mathbf{e})$ is an ex post outcome with $F(\mathbf{e}) > 0$. Denote by $W_i(\mathbf{c}, \mathbf{l})$ the partial derivative of W with respect to c_i. Then, under (A.1)–(A.3) and using the ex post maximization problem (4), $(\mathbf{c}, \mathbf{e})$ must satisfy the following property: if $c_k > 0$, then for all $j = 1, ..., n$,

$$W_k(\mathbf{c}, \mathbf{l}) \geqslant W_j(\mathbf{c}, \mathbf{l}) \tag{16}$$

with equality whenever $c_j > 0$.

We need a slightly stronger implication than (16). Let i be such that $e_i > 0$, and suppose that there is some k with $c_k = 0$. Suppose, further, that there exists $\varepsilon > 0$ such that for all $e'_i \in (e_i - \varepsilon, e_i)$, the ex post consumption vector $c(e'_i, \mathbf{e}_{-i}) \equiv \mathbf{c}'$ has $c'_k > 0$. Then, indeed, we can say that (16) holds for this k *even if* $c_k = 0$. We exclude the verification of this simple observation.

Now, suppose that, contrary to the lemma, this outcome is an equilibrium of the soft game.

Let M be the set of all indices j such that either $c_j > 0$, or with the property that there is $\varepsilon > 0$ such that for all $e'_i \in (e_i - \varepsilon, e_i)$, the ex post consumption vector $c(e'_i, \mathbf{e}_{-i}) \equiv c'$ has $c'_j > 0$. We are going to consider the effect (on i's utility) of a small reduction in e_i by differential methods. By the maximum theorem (and the uniqueness of ex post consumption (given effort)), $\mathbf{c}(\mathbf{e})$ is a continuous function. Therefore the set M is all that counts for the analysis, and for all $k, j \in M$, (16) holds with equality.

Note first that because $(\mathbf{c}, \mathbf{e})$ is an equilibrium and $e_i > 0$, we have $c_i > 0$. Consequently, $i \in M$. Without loss of generality, number the indices in M as $1, ..., m$, and let i be rechristened with the index 1. We now claim that $m \geqslant 2$. Suppose not. Then, because $(\mathbf{c}, \mathbf{e})$ is an equilibrium, we must have

[19] Maskin and Moore [11] study a related problem where the ability of the planner is limited by the possibility that agents may renegotiate the outcome.

$e_j = 0$ for all $j \neq 1$. Therefore, because $F(\mathbf{e}) \geqslant F(\mathbf{e}^*)$ and because we have assumed that the first best output is larger than the output of any individual acting completely on his own, we must have

$$u_c^1(c_1, l_1)\, F_1(\mathbf{e}) < u_l^1(c_1, l_1) \tag{17}$$

Because $M = \{1\}$, it follows from (17) and the definition of M that a small reduction in e_1 will raise 1's welfare, because all other consumptions will continue at zero. This contradicts our supposition that $(\mathbf{c}, \mathbf{e})$ is an equilibrium, and shows that $m \geqslant 2$.

For all $i \in M$, we have the first order condition of the ex post maximization problem: for some $\lambda < 0$,

$$\begin{aligned} W_i(\mathbf{c}, \mathbf{l}) + \lambda &= 0 \\ c_1 + \cdots + c_m &= F(\mathbf{e}). \end{aligned} \tag{18}$$

We are interested in differentiating this system with respect to a parametric change in e_1, and studying dc_1/de_1.[20] Given our remarks above, the differentiation argument will reflect the true story for small changes in e_1, provided that we think of these changes as *reductions*.

Differentiating (18) and defining $W_{ij} \equiv \partial^2 W/\partial c_i \partial c_j$, we obtain the system

$$\begin{pmatrix} 0 & 1 & 1 & \cdots & 1 \\ 1 & W_{11} & W_{12} & \cdots & W_{1m} \\ 1 & W_{21} & W_{22} & \cdots & W_{2m} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & W_{m1} & W_{m2} & \cdots & W_{m} \end{pmatrix} \begin{pmatrix} d\lambda \\ dc_1 \\ dc_2 \\ \vdots \\ dc_m \end{pmatrix} = \begin{pmatrix} F_1(\mathbf{e})\, de_1 \\ \dfrac{\partial W_1(c, \mathbf{l})}{\partial l_i}\, de_1 \\ \dfrac{\partial W_2(c, \mathbf{l})}{\partial l_i}\, de_1 \\ \vdots \\ \dfrac{\partial W_m(c, \mathbf{l})}{\partial l_i}\, de_1 \end{pmatrix} \tag{19}$$

Let us calculate $\partial W_i(\mathbf{c}, \mathbf{l})/\partial l_1$. For $i = 1$,

$$\frac{\partial W_1(\mathbf{c}, \mathbf{l})}{\partial l_1} = \frac{\partial [V_1(\mathbf{u})\, u_c^1(c_1, l_1)]}{\partial l_1} = V_{11}(\mathbf{u})\, u_c^1 u_l^1 + V_1(\mathbf{u})\, u_{cl}^1 \tag{20}$$

[20] (18) defines $(\mathbf{c}, \lambda)$ implicitly as a function of the "parameter" $\mathbf{e}$. Because output is positive, it follows from (A.1)–(A.3) that this function is $\mathscr{C}^1$. Moreover, because the bordered Hessian in (19) has non-zero determinant (see footnote 6), it follows from the implicit function theorem that $\mathbf{c}$ is a differentiable function of $\mathbf{e}$, and in particular of e_1.

For $i \neq 1$, we have

$$\frac{\partial W_i(\mathbf{c}, \mathbf{l})}{\partial l_1} = \frac{\partial [V_i(\mathbf{u})\, u_c^i(c_i, l_i)]}{\partial l_1} = V_{i1}(\mathbf{u})\, u_c^i u_l^1. \tag{21}$$

Next, let us calculate $W_{i1}(\mathbf{c}, \mathbf{l})$. For $i = 1$,

$$W_{11}(\mathbf{c}, \mathbf{l}) = \frac{\partial [V_1(\mathbf{u})\, u_c^1(c_1, l_1)]}{\partial c_1} = V_{11}(\mathbf{u})[u_c^1]^2 + V_1(\mathbf{u})\, u_{cc}^1 \tag{22}$$

while for $i \neq 1$, we see that

$$W_{i1}(\mathbf{c}, \mathbf{l}) = \frac{\partial [V_i(\mathbf{u})\, u_c^i(c_i, l_i)]}{\partial c_1} = V_{i1}(\mathbf{u})\, u_c^1 u_c^i. \tag{23}$$

Now, recalling that leisure is a normal good in individual preferences, we have $u_c^1 u_{cl}^1 - u_l^1 u_{cc}^1 > 0$ (recall (A.2) and footnote 7). Using this and manipulating (20), we have

$$\begin{aligned} \frac{\partial W_1(\mathbf{c}, \mathbf{l})}{\partial l_1} &= \frac{u_l^1}{u_c^1}\left(V_{11}(\mathbf{u})[u_c^1]^2 + V_1(\mathbf{u})\, \frac{u_c^1 u_{cl}^1}{u_l^1} \right) \\ &> \frac{u_l^1}{u_c^1}\, (V_{11}(\mathbf{u})[u_c^1]^2 + V_1(\mathbf{u})\, u_{cc}^1) \\ &= \frac{u_l^1}{u_c^1}\, W_{11}(\mathbf{c}, \mathbf{l}), \end{aligned} \tag{24}$$

using (22), while from (21) and (23) it is easy to see that for all $i \neq 1$,

$$\frac{\partial W_i(\mathbf{c}, \mathbf{l})}{\partial l_1} = \frac{u_l^1}{u_c^1}\, W_{i1}(\mathbf{c}, \mathbf{l}). \tag{25}$$

Now, let us return to (19) and write down the solution for dc_1/de_1. Let U be the determinant of the cofactor of the first 1 (from the left) in the first row of the matrix in (19), and U_i be the determinant of the cofactor of W_{i1}, $i = 1, \ldots, m$. Include in these the signs generated by cofactor expansion. Let U^* be the determinant of the matrix in (19). Then it can be seen that

$$\frac{dc_1}{de_1} = \frac{1}{U^*}\left[F_1(\mathbf{e})U + \frac{\partial W_1(\mathbf{c}, \mathbf{l})}{\partial l_1}\, U_1 + \frac{\partial W_2(\mathbf{c}, \mathbf{l})}{\partial l_1}\, U_2 + \cdots + \frac{\partial W_m(\mathbf{c}, \mathbf{l})}{\partial l_1}\, U_m \right].$$

Recalling from the strict quasiconcavity of $W(\cdot)$ and (18) that U_1 and U^* have different signs, and using (24) and (25), we have

$$\frac{dc_1}{de_1} < \frac{1}{U^*}\left[F_1(\mathbf{e})\, U + \frac{u_l^1}{u_c^1}\,(W_{11} U_1 + W_{21} U_2 + \cdots + W_{m1} U_m) \right] \tag{26}$$

$$= \frac{u_l^1}{u_c^1}. \tag{27}$$

Therefore, using (27),

$$\frac{du^1(c_1, l_1)}{de_1} = u_c^1 \frac{dc_1}{de_1} - u_l^1 < 0, \tag{28}$$

so that (28) proves that a small reduction in e_1 will raise the utility of 1 under the new *ex post* outcome. This contradicts our supposition that $(\mathbf{c}, \mathbf{e})$ is an equilibrium, and completes the proof of the lemma.

Proof of Proposition 1. Observe that the first best allocation is indeed an *ex post* outcome, with positive output, satisfying (15) in Lemma 1. Given the assumption that at the first best, there is i with $(e_i^*, l_i^*) \gg 0$, (15) follows as a necessary condition describing the first best. By Lemma 1, this outcome cannot be an equilibrium of the soft game.

If (A.4) holds, then consider any ex post outcome with $F(\mathbf{e}) \geqslant F(\mathbf{e}^*)$. It follows right away from (A.4) that (15) is satisfied. By Lemma 1, this outcome cannot be an equilibrium of the soft game.

Proof of Proposition 2. For any outcome $(\mathbf{c}, \mathbf{e})$, define its *Rawlsian value* as $R(\mathbf{c}, \mathbf{e}) \equiv \min_i u^i(c_i, l_i)$. Now note that under the first best outcome, we must have, for *every* i,

$$R(\mathbf{c}^*, \mathbf{e}^*) = u^i(c_i^*, l_i^*). \tag{29}$$

The reason is that for each i, either $c_i^* > 0$ or $l_i^* > 0$ (or both), because $u^i(c_i^*, l_i^*) \geqslant u^i(0, L_i)$. Therefore, if (29) were not true, we could always improve the Rawlsian value of the outcome by small changes.

Given this claim, we first prove that the first best outcome is indeed an equilibrium. Suppose not. Then for some i, there is $e_i \in [0, L_i]$ such that

$$u^i(c_i, l_i) > u^i(c_i^*, l_i^*), \tag{30}$$

where c_i is the ith component of $\mathbf{c} \equiv \mathbf{c}(\mathbf{e})$, and $\mathbf{e} \equiv (e_i, \mathbf{e}^*_{-i})$.

Because $u^i(c_i^*, l_i^*) \geqslant u^i(0, L_i)$ and because (30) holds, it must be the case that $c_i > 0$. It follows that

$$R(\mathbf{c}, \mathbf{e}) = u^i(c_i, l_i) \tag{31}$$

For if not, the Rawlsian value of $(\mathbf{c}, \mathbf{e})$ can be improved by only changing the consumption allocation, a contradiction to the fact that $(\mathbf{c}, \mathbf{e})$ is an *ex post* outcome (use the fact that $c_i > 0$).

Combining (29), (30), and (31), we have

$$R(\mathbf{c}, \mathbf{e}) > R(\mathbf{c}^*, \mathbf{e}^*),$$

but this contradicts the fact that (c^*, e^*) is the Rawlsian first best.

Now we prove the second part of the proposition. First, we show that if $(\mathbf{c}, \mathbf{e})$ is an equilibrium, then for all $i = 1, ..., n$,

$$R(\mathbf{c}, \mathbf{e}) = u^i(c_i, l_i). \tag{32}$$

Suppose this is not true. Then for some pair i, j, $u^i(c_i, l_i) > u^j(c_j, l_j)$. But then it must be the case (by an earlier argument) that $c_i = 0$. Consequently, $u^j(c_j, l_j) < u^i(0, l_i) \leqslant u^i(0, L_i) = u^j(0, L_j)$, where the last equality follows from normalization. But then $(\mathbf{c}, \mathbf{e})$ cannot be an equilibrium, for j can guarantee himself at least $u^j(0, L_j)$ by deviating.

With this established, we return to the main proof. Suppose that the proposition is false. Then there exists an equilibrium $(\mathbf{c}, \mathbf{e}) \neq (\mathbf{c}^*, \mathbf{e}^*)$, such that at least one player gets strictly more utility than inaction. By virtue of (32) and assumption (R), it follows that *every* player gets strictly more than inaction. This observation will be used below.

Because the first best is unique, $R(\mathbf{c}^*, \mathbf{e}^*) > R(\mathbf{c}, \mathbf{e})$. Define $\mathbf{e}(t) \equiv t\mathbf{e} + (1-t)\,\mathbf{e}^*$ for $t \in [0, 1]$, and $\mathbf{c}(t) \equiv \mathbf{c}(\mathbf{e}(t))$. Then it is easy to check that $R(\mathbf{c}(t), \mathbf{e}(t)) > R(\mathbf{c}, \mathbf{e})$ for all $t \in (0, 1)$ (use (A.1), (A.2), (29) and (32)). Using (32) again, it follows that for all $t \in (0, 1)$ and all i,

$$u^i(c_i(t), l_i(t)) > u^i(c_i, l_i) \tag{33}$$

Now we claim that the following is true:

There exists j such that *either* (i) $e_j = L_j$ and $u^j_c(c_j, l_j)\, F_j(\mathbf{e}) < u^j_l(c_j, l_j)$, or (ii) $e_j \in (0, L_j)$ and $u^j_c(c_j, l_j)\, F_j(\mathbf{e}) \neq u^j_l(c_j, l_j)$, or (iii) $e_j = 0$ and $u^j_c(c_j, l_j)\, F_j(\mathbf{e}) > u^j_l(c_j, l_j)$.

To prove the claim, first recall that each player gets strictly more than inaction at $(\mathbf{c}, \mathbf{e})$, so that by (A.2), $(u^i_c(c_i, l_i), u^i_l(c_i, l_i)) \gg 0$ for all i. Using (A.2) again, together with (33), we see that for all i,

$$(c_i(t) - c_i) + (l_i(t) - l_i)\,\frac{u^i_l}{u^i_c} > 0$$

or equivalently, for all i,

$$(c_i(t) - c_i) - (e_i(t) - e_i)\,\frac{u^i_l}{u^i_c} > 0. \tag{34}$$

Now, if the claim is false, then, noting that $e_i(t) \in [0, L_i]$ for all i, we can use (34) to deduce that

$$(c_i(t) - c_i) - (e_i(t) - e_i)\, F_i(\mathbf{e}) > 0.$$

Summing this inequality over all i, we have

$$F(\mathbf{e}(t)) - F(\mathbf{e}) > (e_1(t) - e_1)\, F_1(\mathbf{e}) + \cdots + (e_n(t) - e_n)\, F_n(\mathbf{e}). \tag{35}$$

Now observe that because $(\mathbf{c}, \mathbf{e})$ provides more utility than inaction, $F(\mathbf{e}) > 0$. But then (35) contradicts (A.1), which assumes that F is differentiable whenever $F(\mathbf{e}) > 0$, and that F is concave. This establishes the claim.

Pick j as given by the claim. If part (i) of the claim is true, notice that because $(\mathbf{c}, \mathbf{e})$ is an equilibrium, $c_j > 0$. Then, a small reduction in e_j makes j *strictly* better off even if j pays for the entire reduction in output from his own consumption. This improvement in j's utility persists even if j must pay a small additional amount to each of the other agents. This proves that the Rawlsian value of $(\mathbf{c}, \mathbf{e})$ can be improved by a *unilateral* change made by j, by going to an ex post outcome $(\mathbf{c}', \mathbf{e}')$ (where $\mathbf{e}'$ differs from $\mathbf{e}$ only in the jth component). That is, $R(\mathbf{c}', \mathbf{e}') > R(\mathbf{c}, \mathbf{e})$. But note that, as a result, $u^j(c_j', l_j') \geqslant R(\mathbf{c}', \mathbf{e}') > R(\mathbf{c}, \mathbf{e}) = u^j(c_j, l_j)$ (where the last equality uses (32)). This contradicts the supposition that $(\mathbf{c}, \mathbf{e})$ is an equilibrium.

Finally, if part (ii) or (iii) of the claim is true, use the same argument as above, if $u_c^j(c_j, l_j)\, F_j(\mathbf{e}) < u_l^j(c_j, l_j)$, or its reverse (increase e_j), if $u_c^j(c_j, l_j)\, F_j(\mathbf{e}) > u_l^j(c_j, l_j)$, to arrive at a contradiction.

This completes the proof of the proposition. ▮

Proof of Proposition 3. By our assumptions, the first best is symmetric and has

$$0 < e^* < L.$$

It follows from the first-order conditions characterizing the first best that

$$u_c(c^*, L - e^*)\, F_i(e^*, ..., e^*) = u_l(c^*, L - e^*),$$

and for every $e > e^*$ and $c \equiv F(e, ..., e)/n$,

$$u_c(c, L - e)\, F_i(e, ..., e) \leqslant u_l(c, L - e).$$

Moreover, these hold for all i. It follows from Lemma 1 that no symmetric $e \geqslant e^*$ can be an equilibrium of the soft game.

To prove the remainder of the proposition, fix a common utility level $\bar{u}$ for all but the first individual, and let u_1 denote the utility level of individual 1. Then by applying the characterization of increased egalitarianism for smooth welfare functions, we see that

$$\bar{u} \geqslant u_1 \Leftrightarrow \frac{V'_2}{V'_1} \leqslant \frac{V_2}{V_1}, \tag{36}$$

where a subscript 1 denotes the marginal social welfare contributed by the first individual's utility, and the subscript 2 is the corresponding contribution by the others. Define a "utility possibility frontier" (given efforts) in the modified soft game by

$$\Psi(\bar{u}, e_1, e) \equiv u(F(e_1, e, ..., e) - (n-1)\, c, L - e_1), \tag{37}$$

where c is chosen such that $\bar{u} = u(c, L-e)$. Ψ is downward sloping and concave with respect to $\bar{u}$, by (A.5).

The social planner chooses consumption to maximize V on Ψ. Let us introduce some notation. We will denote by $U_1(e_1, V)$ the utility of agent 1 when the action taken by him is e_1, the welfare function is V, the action of the remaining agents is fixed at e (which is not explicitly carried in the notation), and the planner chooses *ex-post* consumption optimally. Likewise, we will denote by $U(e_1, V)$ the utility of each of the other agents under exactly the same state of affairs.[21] We make two observations. First,

$$c_1(e, e; V) = c(e, e; V) = \frac{1}{n} F(e, ..., e) \tag{38}$$

for any V. Second, it is possible to show, using the concavity of Ψ, the strict quasiconcavity of V and V', and (36),[22] that

$$U_1(e_1, V) \geqslant U(e_1, V) \Leftrightarrow U_1(e_1, V) \geqslant U_1(e_1, V') \geqslant U(e_1, V') \geqslant U(e_1, V), \tag{39}$$

when V' is at least as egalitarian as V. Moreover, precisely the opposite chain of inequalities hold on the RHS of (39) if the opposite inequality holds on the LHS of (39).

Let V' be at least as egalitarian than V. We claim that

[I] If $e \in B(e; V)$, then $e \leqslant \max B(e; V')$.

Suppose not. Then $\max B(e; V') < e$. So there exists $x \in [0, e)$ such that

$$U_1(x, V') > U_1(e, V'). \tag{40}$$

[21] Thus $U_1(e_1, V) = u(c_1(e_1, e, V), L - e_1)$ and $U(e_1, V) = u(c(e_1, e, V), L - e)$.

[22] In fact, this is the only point at which (36) is used in the proof.

Because $e \in B(e, V)$, it must be the case that

$$U_1(x, V) \leqslant U_1(e, V). \tag{41}$$

Noting from (38) that $U_1(e, V) = U_1(e, V')$, we may combine (40) and (41) to obtain

$$U_1(x, V) < U_1(x, V'). \tag{42}$$

Combine (42) with (39) and the claim immediately following (39). We may deduce that

$$U_1(x, V') \leqslant U(x, V'). \tag{43}$$

Noting from (38) that $U_1(e, V') = U(e, V')$, and combining this observation with (40) and (43),

$$U(x, V') \geqslant U_1(x, V') > U_1(e, V') = U(e, V'). \tag{44}$$

In words, (40) tells us that under V', player 1 is better off choosing x rather than e. At the same time, (44) tells us that the rest of the players are *also* better off when player 1 chooses x instead of e. Thus a vector-inferior collection of efforts $(x, e, ..., e)$ instead of $(e, ..., e)$ (which in turn is lower than the first best $(e^*, ..., e^*)$) leads to a Pareto-improvement. Under our assumptions, this cannot be.

To see this more formally, note that if $x < e < e^*$, then

$$U_1(e, V') = U_1(e, V) > \Psi(\bar{u}, x, e) \tag{45}$$

as long as $\Psi(\bar{u}, x, e) \leqslant \bar{u}$. By putting $\bar{u} = U(x, V')$, $\psi(\bar{u}, x, e) = U_1(x, V')$, and using (44), we contradict (45). This completes the proof of Claim [I].

Denote the maximal effort level among all symmetric equilibria under V' by $\bar{e}$. We have already shown that $\bar{e} < e^*$. We claim that

[II] $\max B(e; V') < e$ for all $e \in (\bar{e}, L]$.

To see this, first observe that $\max B(L, V') \leqslant L$ simply by definition. Now suppose that the claim is false. Then for some $e \in (\bar{e}, L]$, $\max B(e; V') \geqslant e$. Moreover, $B(., V')$ is convex-valued and upperhemicontinuous. But this establishes (using a simple argument analogous to the Intermediate Value Theorem) the existence of some $e' > \bar{e}$ such that $e' \in B(e'; V')$. Moreover, a fixed point of $B(, ; V')$ corresponds to a symmetric equilibrium. But this contradicts the definition of $\bar{e}$ as the *largest* symmetric equilibrium effort level under V'.

To complete the proof of the proposition, suppose that there exists $e \in (\bar{e}, L]$ which belongs to $B(e; V)$. Then $\max B(e; V') \geqslant e > \max B(e; V')$, where the first inequality follows from Claim [I], and the second inequality follows from Claim [II]. This is a contradiction, and the proof is complete.

REFERENCES

1. G. S. Becker, A theory of social interactions, *J. Polit. Econ.* **82** (1974), 1063–1094.
2. T. C. Bergstrom, A fresh look at the rotten kid theorem—and other household mysteries, *J. Polit. Econ.* **97** (1989), 1138–1159.
3. B. D. Bernheim and O. Stark, Altruism within the family reconsidered: Do nice guys finish last?, *Amer. Econ. Rev.* **78** (1988), 1034–1045.
4. P. Dasgupta and D. Ray, Inequality as a determinant of malnutrition and unemployment: Policy, *Economic J.* **97** (1987), 177–188.
5. B. Dutta and D. Ray, A concept of egalïtarianism under participation constraints, *Econometrica* **57** (1989), 615–635.
6. R. H. Frank, If *homo economicus* could choose his own utility function, would he want one with a conscience?, *Amer. Econ. Rev.* **77** (1987), 593–604.
7. J. Geanakoplos and B. Nalebuff, "On a Fundamental Conflict between Equity and Efficiency," Economics Discussion Paper No. 137, Woodrow Wilson School, Princeton University, 1988.
8. B. Holmstrom, Moral hazard in teams, *Bell J. Econ.* **13** (1982), 324–340.
9. A. Lindbeck and J. W. Weibull, Altruism and time consistency: The economics of fait accompli, *J. Polit. Econ.* **96** (1988), 1165–1182.
10. G. C. Loury, Intergenerational transfers and the distribution of earnings, *Econometrica* **49** (1981), 843–867.
11. E. Maskin and J. Moore, Implementation and renegotiation, mimeo, 1990.
12. J. Mirrlees, A pure theory of underdeveloped economies, *in* "Agriculture in Development Theory" (L. Reynolds, Ed.), Yale Univ. Press, New Haven, CT, 1975.
13. H. Moulin, "Axioms of Cooperative Decision-Making," Cambridge Univ. Press, Cambridge, UK, 1988.
14. A. M. Okun, "Equality and Efficiency: The Big Tradeoff," Brookings Institution, Washington, DC, 1975.
15. K. Sydsaeter, Letter to the editor on some frequently occurring errors in the economic literature concerning problems of maxima and minima, *J. Econ. Theory* **9** (1974), 464–466.