Common Agency and Coordination: General Theory and Application to Tax Policy[*]

by
Avinash Dixit, Princeton University
Gene M. Grossman, Princeton University
and
Elhanan Helpman, Tel Aviv University and CIAR

December 14, 1999

Abstract

We develop a model of common agency with complete information and general preferences with non-transferable utility, and prove that the principals' Nash equilibrium in truthful strategies implements an efficient action. We apply this theory to construct a positive model of public finance, where organized special interests can lobby the government for consumer and producer taxes or subsidies and targeted lump-sum taxes or transfers. The lobbies use only the non-distorting transfers in their non-cooperative equilibrium, but their inter-group competitition for transfers turns into a prisoners' dilemma in which the government captures all the gain that is potentially available to the parties. Therefore we suggest that pressure groups capable of sustaining an ex ante agreement will make a commitment to forgo direct transfers and to confine their lobbying to distorting taxes and subsidies.

Addresses of authors:

Avinash Dixit, Department of Economics, Princeton University, Princeton, NJ 08544, USA. E-mail: dixitak@princeton.edu

Gene M. Grossman, Woodrow Wilson School, Princeton University, Princeton, NJ 08544, USA. E-mail: grossman@wws.princeton.edu

Elhanan Helpman, Eitan Berglas School of Economics, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: helpman@econ.tau.ac.il

---

1   Introduction

Common agency is the name given to a multilateral relationship where several principals simultaneously try to influence the actions of an agent. Such situations occur very frequently, particularly in the political processes that generate economic policy. For example, legislatures respond to many diverse pressures, including those from voters, contributors, and party officials. Administrative agencies are formally responsible to either the legislative or the executive branches, but are in practice influenced by the courts, the media, and various interest and advocacy groups. In the European Union, several sovereign governments deal with a common policy-making apparatus in Brussels. In the United States, growing decentralization of economic power to the states and localities may give governments at these levels something of the standing of principals in relation to the federal government.

Information asymmetries can be important in a common agency just as in an agency with a single principal. However, even with complete information, the existence of multiple principals introduces the new issues of whether they can achieve an outcome that is efficient for the group of players (the principals and the agent together), and of how the available economic surplus gets distributed among players. Bernheim and Whinston (1986) show that the non-cooperative game among the principals does have an efficient equilibrium. Their model has found many applications, including models of lobbying for tariffs (Grossman and Helpman, 1994) and for consumer and producer taxes and subsidies (Dixit, 1995).

However, the Bernheim-Whinston model assumes quasi-linear preferences, so monetary transfers are equivalent to transferable utility among the principals and their common agent. This limits the applications of their model in two ways. First, the agent's actions become independent of the distribution of payoffs among the players.[1] In political settings, this means that the policy-setting authority cannot care about income distribution *per se.* Second, since transfers between players do not have diminishing marginal payoffs, equilibria may not exist, or may exist but have unrealistically large transfers and extreme distributional implications. Dixit (1995) discusses this feature of the model in the context of tax policy.

In most economic applications, money is transferable, but the players' payoffs are not linear in money. In this paper we generalize the theory of common agency to handle such situations. We thereby hope greatly to enlarge the scope of applicability of the theory.

---

[1]Note the parallel with the Coase theorem, where under quasi-linear utility (no income effects on the activities in question), resource allocation is independent of distribution.

When utility is not directly transferable across players, efficiency of equilibrium is not immediately obvious and needs careful and rigorous proof. We therefore begin by showing that the agent's actions in equilibrium still achieve efficiency for the group of players (Theorem 2). Of course, the actions are no longer independent of the distribution of payoffs among the principals (or between the principals and the agent), and in equilibrium the two sets of magnitudes must be determined simultaneously.[2]

We then consider a political process of economic policy-making in the common agency framework. A subset of all individuals are allowed to lobby the government, and promise contributions in return for policy choices. The government cares for social welfare defined over the utilities of all individuals (lobbying or not) and for its receipts from the lobbyists. The efficiency theorem then says that the government uses the available instruments in a Pareto efficient manner.

We develop this in greater detail in a positive model of the formation of tax policy that is rich enough to become a counterpart to the familiar normative model of Diamond and Mirrlees (1971). The policy instruments are commodity tax or subsidy policies and individualized lump-sum transfers, and the political process is of the form of lobbying described above. We show how the strict concavity of utilities in money incomes makes determinate the levels of transfers in the political equilibrium.

Here the efficiency result implies that only the non-distorting lump-sum transfers are used in the political equilibrium, not consumption or production taxes or subsidies. However, this should not be interpreted as a general proof of efficiency of politics. The game of lobbying for transfers turns into a prisoners' dilemma for the lobbyists. Indeed, under mild additional assumptions, we find that the government captures all the gains that exist in the common agency relationship.[3] This suggests that if the lobbyists could commit ex ante to a "constitution" for lobbying, they would all agree not to lobby for lump-sum transfers. This opens the way for the use of economically inferior instruments such as production subsidies,

---

[2]The parallel with the Coase Theorem or the core with non-transferable utility should again be apparent. However, we should stress that ours is an equilibrium of a non-cooperative game, not a cooperative solution concept.

[3]Grossman and Helpman (1994) found a similar tendency for the lobbies' gains to be lost when tariffs were replaced as instruments by the economically more efficient production subsidies; this result with lump-sum transfers is even more dramatic and extreme.

with an attendant violation of production efficiency in the political equilibrium, contrary to an important general feature of the normative optimum (Diamond and Mirrlees, 1971).

2   General Theory

Consider the following problem. There is a set L of principals. For each $i \in L$, principal i has continuous preferences $W_i(\mathbf{a}, c_i)$, where the vector $\mathbf{a}$ denotes the agent's action and the scalar c_i denotes principal i's payment to the agent. Each principal's preference function is declining in his payment to the agent. The agent's continuous preference function is $G(\mathbf{a}, \mathbf{c})$, where $\mathbf{c}$ is the L-dimensional vector of the principals' payments. The function G is increasing in each component of $\mathbf{c}$. Thus, for any given action, each principal dislikes making contributions and the agent likes receiving them; their preferences with regard to actions are not restricted in general, but we will place some specific restrictions for particular results below. We refer to the values of the functions $W_i(\mathbf{a}, c_i)$ and $G(\mathbf{a}, \mathbf{c})$ as the utility levels of the principals and the agent respectively.

Principal i can choose a payment function $C_i(\mathbf{a})$ from a set $\mathcal{C}_i$ and the agent can choose $\mathbf{a}$ from a set $\mathcal{A}$. The sets $\mathcal{C}_i$ and $\mathcal{A}$ describe feasibility and institutional constraints. For example, from feasibility considerations $\mathcal{C}_i$ may consist only of functions that provide principal i with a non-negative income. Or it may consist only of non-negative functions, implying that the principal can pay the agent but not the reverse. This would describe an institutional constraint. And it may contain only functions with an upper bound on payments, thereby describing another institutional constraint. Similarly, $\mathcal{A}$ may describe institutional or feasibility constraints on the actions of the agent. If, for example, an element of $\mathbf{a}$ equals one plus an ad-valorem tax rate, then feasibility requires $\mathcal{A}$ to contain only non-negative vectors.

Throughout, we maintain the following assumption on the sets of feasible payment functions:

Assumption 1: Let $C_i \in \mathcal{C}_i$. Then $C_i(\mathbf{a}) \geq 0$ for all $\mathbf{a} \in \mathcal{A}$; and every payment function C_i^* that satisfies: (i) $C_i^*(\mathbf{a}) \geq 0$ for all $\mathbf{a} \in \mathcal{A}$, and (ii) $C_i^*(\mathbf{a}) \leq C_i(\mathbf{a})$ for all $\mathbf{a} \in \mathcal{A}$, also belongs to $\mathcal{C}_i$.

Explanation: Payments from the principals to the agent have to be nonnegative, and if a payment function is feasible, all "smaller" payment functions are also feasible. This conforms to the requirements of all relevant economic applications.

Equilibrium

Our aim is to construct and study a concept of equilibrium for a two-stage game. In the second stage, the agent chooses an action optimally, given the payment functions of all the principals. In the first stage, each principal chooses a payment schedule, knowing that all the other principals are simultaneously and non-cooperatively choosing their own payment schedules, and looking ahead to the response of the agent in the second stage.

We will denote magnitudes pertaining to an equilibrium by the superscript $\circ$. Since the game is non-cooperative, we will have to start with a "candidate" for such an equilibrium, and study the consequences of allowing the strategies to deviate from this, one player at a time. For this purpose we establish the following notation: $\mathbf{C}^\circ(\mathbf{a})$ will denote the vector of contribution functions with component functions $C_j^\circ(\mathbf{a})$, for all $j \in L$; while $(\{C_j^\circ(\mathbf{a})\}_{j \neq i}, c)$ will denote the vector where the i-th component is replaced by c, and all the other components $j \neq i$ are held fixed at $C_j^\circ(\mathbf{a})$. Sometimes c itself may be the value of another payment function $C_i(\mathbf{a})$ for principal i.

We begin by defining the principals' best response strategies.

Definition 1: A payment function $C_i^\circ \in \mathcal{C}_i$ is a best response of principal i to the payment functions $\{C_j^\circ\}_{j \in L, j \neq i}$ of the other principals if there does not exist a payment function $C_i \in \mathcal{C}_i$ such that

(i) $W_i[\mathbf{a}_i, C_i(\mathbf{a}_i)] > W_i[\mathbf{a}^\circ, C_i^\circ(\mathbf{a}^\circ)]$; where

(ii) $\mathbf{a}_i = \arg\max_{\mathbf{a} \in \mathcal{A}} G[\mathbf{a}, (\{C_j^\circ(\mathbf{a})\}_{j \neq i}, C_i(\mathbf{a}))]$, and

(iii) $\mathbf{a}^\circ = \arg\max_{\mathbf{a} \in \mathcal{A}} G[\mathbf{a}, \mathbf{C}^\circ(\mathbf{a})]$.

Explanation: The best response calculation of principal i holds fixed the simultaneously chosen strategies (payment functions) of all the other principals at their candidate equilibrium positions, but recognizes that in the second stage the agent will optimize with respect to these payment functions along with any deviated function proposed by principal i. If principal i cannot find another feasible payment function that yields a better outcome for him, taking into account the agent's anticipated response, then the original candidate payment function C_i° is principal i's best response to the candidate functions C_j° of all the other principals.

Next we define equilibrium. This is the standard definition of a subgame perfect Nash equilibrium for this two-stage game; it is stated explicitly only so that we can refer to the specific conditions (a)-(c) later.

Definition 2: An equilibrium of the common agency problem consists of a vector of payment functions $\mathbf{C}^\circ = \{C_i^\circ\}_{i \in L}$ and a policy vector $\mathbf{a}^\circ$ such that:

(a) $C_i^\circ \in \mathcal{C}_i$ for all $i \in L$;

(b) $\mathbf{a}^\circ = \arg\max_{\mathbf{a} \in \mathcal{A}} G[\mathbf{a}, \mathbf{C}^\circ(\mathbf{a})]$; and

(c) for every $i \in L$ the payment function C_i° is a best response of principal i to the payment functions $\{C_j^\circ\}_{j \in L, j \neq i}$ of the other principals.

The following theorem provides a characterization of an equilibrium:

Theorem 1: A vector of payment functions $\mathbf{C}^\circ = \{C_i^\circ\}_{i \in L}$ and a policy vector $\mathbf{a}^\circ$ constitute an equilibrium if and only if:

(a) $C_i^\circ \in \mathcal{C}_i$ for all $i \in L$;

(b) $\mathbf{a}^\circ = \arg\max_{\mathbf{a} \in \mathcal{A}} G[\mathbf{a}, \mathbf{C}^\circ(\mathbf{a})]$; and

(c) for every $i \in L$:

$$[\mathbf{a}^\circ, C_i^\circ(\mathbf{a}^\circ)] = \arg\max_{(\mathbf{a}, c)} W_i(\mathbf{a}, c), \tag{2.1}$$

subject to $\mathbf{a} \in \mathcal{A}$, $c = C_i(\mathbf{a})$ for some $C_i \in \mathcal{C}_i$, and

$$G[\mathbf{a}, (\{C_j^\circ(\mathbf{a})\}_{j \neq i}, c)] \geq \max_{\mathbf{a}' \in \mathcal{A}} G[\mathbf{a}', (\{C_j^\circ(\mathbf{a}')\}_{j \neq i}, 0)]. \tag{2.2}$$

Explanation: Observe that (a) and (b) are mere restatements of the corresponding parts of Definition 2. The reformulation of (c) is the key aspect of Theorem 1; it focuses on the relationship between the agent and one of the principals, and helps determine how the potential gains from this relationship get allocated between them in equilibrium.

Examine the situation from the perspective of principal i. He takes as given the strategies of all other principals $j \neq i$, and contemplates his own choice. He must provide the agent

at least the level of utility which the agent could get from his outside option, namely by choosing his best response to the payment functions offered by all the other principals when principal i offers nothing. This is what constraint (2.2) expresses. Subject to this constraint, principal i can propose to the agent an action and a feasible payment that maximizes his own utility. That is the content of equation (2.1). Then Theorem 1 says that such constrained maximization by each principal is equivalent to equilibrium as previously defined.

The intuition behind our result can be appreciated with the aid of Figure 1. Suppose for the sake of illustration that the action is a scalar. Curve $G_i G_i$ depicts combinations of the action a (on the horizontal axis) and payments c by principal i (on the vertical axis) that give the agent a fixed level of utility when the contribution functions of the other agents are given. The particular indifference curve shown in the figure describes the highest welfare the agent can attain when principal i makes no contribution whatsoever (his payment function coincides with the horizontal axis); the agent then chooses the action associated with the point labelled A_{-i}. The shaded rectangle depicts the combinations of feasible actions and feasible payment levels (there is an upper bound on payments, payments have to be non-negative, and the action is bounded below and above). Considering the agent's option to take action A_{-i}, the best the principal i can do is to design a payment schedule that induces the agent to choose a point in the shaded area that lies above or on the indifference curve $G_i G_i$. Suppose the principal's welfare is increasing in the action. Then his indifference curves are upwards sloping. In the event, he will choose a point on the rising portion of $G_i G_i$ that is both feasible and gives him the highest welfare level, namely the tangency point A between his indifference curve $W_i W_i$ and $G_i G_i$. It is easy to see from the figure how the agent can construct a payment schedule that induces the agent to choose point A. For example, he might offer a schedule that coincides with the horizontal axis until some point to the right of A_{-i}, and then rises to a tangency with $G_i G_i$ at A without ever crossing that curve.

Proof of Theorem 1: Conditions (a) and (b) of the theorem just restate those in the definition of an equilibrium. It therefore remains to prove necessity and sufficiency of condition (c) given that (a) and (b) hold. We prove both parts "negatively," that is, start by assuming that condition (c) in one place (definition or theorem) is violated and proving that this implies condition (c) must be violated in the other place as well.

Necessity: Suppose that condition (c) of the theorem does not hold for some $i \in L$. Then there exists a vector $(\mathbf{a}^*, c^*)$ that solves the maximization problem on the right hand side of

7

equation (2.1) when applied to that i, and yields $W_i(\mathbf{a}^*, c^*) > W_i[\mathbf{a}^\circ, C_i^\circ(\mathbf{a})]$. Since $(\mathbf{a}^*, c^*)$ satisfies the constraints of the maximization problem, there exists a payment function $\hat{C}_i \in \mathcal{C}_i$ with $\hat{C}_i(\mathbf{a}^*) = c^*$. Now define the function $\varphi_i(\mathbf{a})$ that satisfies

$$G[\mathbf{a}, (\{C_j^\circ(\mathbf{a})\}_{j \neq i}, \varphi_i(\mathbf{a}))] = G[\mathbf{a}^*, (\{C_j^\circ(\mathbf{a}^*)\}_{j \neq i}, c^*)] \qquad \text{for all } \mathbf{a} \in \mathcal{A}.$$

Since the function $G(\mathbf{a}, \mathbf{c})$ is assumed to be increasing in each component of $\mathbf{c}$, φ_i is a well defined function and $\varphi_i(\mathbf{a}^*) = c^*$. Moreover, $(\mathbf{a}^*, c^*)$ satisfies (2.2). Combining this with the definition of φ_i, we have, for all $\mathbf{a} \in \mathcal{A}$,

$$
\begin{aligned}
G[\mathbf{a}, (\{C_j^\circ(\mathbf{a})\}_{j \neq i}, \varphi_i(\mathbf{a}))] &= G[\mathbf{a}^*, (\{C_j^\circ(\mathbf{a}^*)\}_{j \neq i}, c^*)] \\
&\geq \max_{\mathbf{a}' \in \mathcal{A}} G[\mathbf{a}', (\{C_j^\circ(\mathbf{a}')\}_{j \neq i}, 0)] \\
&\geq G[\mathbf{a}, (\{C_j^\circ(\mathbf{a})\}_{j \neq i}, 0)].
\end{aligned}
$$

The first step is the definition of φ_i, the second is the constraint (2.2) satisfied by $(\mathbf{a}^*, c^*)$, and the third follows from just a particular choice of $\mathbf{a}'$ in this constraint.

Since $G(\mathbf{a}, \mathbf{c})$ is increasing in each component of $\mathbf{c}$, this implies that $\varphi_i(\mathbf{a}) \geq 0$ for all $\mathbf{a} \in \mathcal{A}$. It follows from Assumption 1 that the payment function C_i^* defined by $C_i^*(\mathbf{a}) = \min[\hat{C}_i(\mathbf{a}), \varphi_i(\mathbf{a})]$ for all $\mathbf{a} \in \mathcal{A}$ is feasible for principal i, and that $C_i^*(\mathbf{a}^*) = c_i^*$. Therefore, for all $\mathbf{a} \in \mathcal{A}$,

$$
\begin{aligned}
G[\mathbf{a}, (\{C_j^\circ(\mathbf{a})\}_{j \neq i}, C_i^*(\mathbf{a}))] &\leq G[\mathbf{a}, (\{C_j^\circ(\mathbf{a})\}_{j \neq i}, \varphi_i(\mathbf{a}))] \\
&= G[\mathbf{a}^*, (\{C_j^\circ(\mathbf{a}^*)\}_{j \neq i}, c_i^*)] \\
&= G[\mathbf{a}^*, (\{C_j^\circ(\mathbf{a}^*)\}_{j \neq i}, C_i^*(\mathbf{a}^*))].
\end{aligned}
$$

The first step follows from $C_i^*(\mathbf{a}) \leq \varphi_i(\mathbf{a})$, the second is the definition of φ_i, and the third follows from $C_i^*(\mathbf{a}^*) = c_i^*$. This chain proves that

$$\mathbf{a}^* = \arg\max_{\mathbf{a} \in \mathcal{A}} G[\mathbf{a}, (\{C_j^\circ(\mathbf{a})\}_{j \neq i}, C_i^*(\mathbf{a}))].$$

But $W_i[\mathbf{a}^*, C_i^*(\mathbf{a}^*)] > W_i[\mathbf{a}^\circ, C_i^\circ(\mathbf{a}^\circ)]$; which violates condition (c) in the definition of an equilibrium.

Sufficiency: Suppose that condition (c) in the definition of an equilibrium does not hold for some i. Namely, suppose that there exists a feasible payment function C_i and a feasible action $\mathbf{a}_i$ such that:

(i) $\mathbf{a}_i = \arg\max_{\mathbf{a}\in\mathcal{A}} G[\mathbf{a}, (\{C_j^\circ(\mathbf{a})\}_{j\neq i}, C_i(\mathbf{a}))\,]$ and

(ii) $W_i[\mathbf{a}_i, C_i(\mathbf{a}_i)] > W_i[\mathbf{a}^\circ, C_i^\circ(\mathbf{a}^\circ)]$.

Then it follows from Assumption 1 that $C_i(\mathbf{a}_i) \geq 0$, which together with (i) above implies that $[\mathbf{a}_i, C_i(\mathbf{a}_i)]$ is feasible in the maximization problem in condition (c) of Theorem 1. Together with (ii) above, this contradicts condition (c) of the theorem; namely, $[\mathbf{a}^\circ, C_i^\circ(\mathbf{a}^\circ)]$ does not solve the maximum problem in condition (c) of the theorem. □

Corollary 1: Let $(\mathbf{C}^\circ, \mathbf{a}^\circ)$ be an equilibrium. Then, for each $i \in L$,

$$G[\mathbf{a}^\circ, \mathbf{C}^\circ(\mathbf{a}^\circ)] = \max_{\mathbf{a}\in\mathcal{A}} G[\mathbf{a}, (\{C_j^\circ(\mathbf{a})\}_{j\neq i}, 0)\,].$$

Explanation: This says that the utility level of the agent in equilibrium is the same as what he would get if any one of the principals were to contribute zero while all the others maintained their equilibrium payment *functions*, and the agent then were to choose his optimum action in response to this deviation. The intuition is implicit in our discussion of condition (c) of the theorem. Each principal must ensure that the agent gets a utility equal to his outside opportunity; it is not in the principal's interest to give the agent any more. The formal proof also closely follows the intuition.

Proof of Corollary 1: This result follows directly from conditions (b) and (c) of Theorem 1. There are two possibilities: $C_i^\circ(\mathbf{a}^\circ)$ is either $= 0$ or > 0. In the former case the corollary is obvious. In the latter case if the inequality in the constraint (2.2) of the theorem's condition (c) is strict, then c can be further reduced without altering $\mathbf{a}$ and still preserve feasibility. This raises $W_i(\mathbf{a}, c)$, thereby violating this condition. □

Truthful Equilibria

The above model can have multiple subgame perfect Nash equilibria, some of which can be inefficient. As in Bernheim and Whinston (1986), we now develop a refinement that selects equilibria that implement Pareto efficient actions (the concept of Pareto efficiency is of course constrained by the set of available actions). We first establish a closely related property that is necessary for all equilibria given some additional restrictions, and then develop the property that is sufficient for efficient equilibria.

Proposition 1: *Local Truthfulness* Let the preference functions $(\{W_i\}_{i \in L}, G)$ be differentiable and let the sets of feasible payment schedules be restricted to functions that are differentiable where positive. Then whenever the equilibrium is interior, in the sense that $\mathbf{a}^\circ$ belongs to the interior of $\mathcal{A}$ and $C_i^\circ(\mathbf{a}^\circ) > 0$ for all $i \in L$, then the equilibrium payment functions exhibit *local truthfulness* at the equilibrium point, in the sense that

$$\nabla_{\mathbf{a}} C_i^\circ(\mathbf{a}^\circ) = -\left(\frac{\partial W_i[\mathbf{a}^\circ, C_i^\circ(\mathbf{a}^\circ)]}{\partial c_i} \right)^{-1} \nabla_{\mathbf{a}} W_i[\mathbf{a}^\circ, C_i^\circ(\mathbf{a}^\circ)] \text{ for all } i \in L, \qquad (2.3)$$

where the operator ∇ applied to a function denotes the gradient vector of the partial derivatives of the function with respect to the vector argument which appears as the subscript of the operator.

Explanation: By the implicit function theorem, the right hand side is just the gradient of c with respect to $\mathbf{a}$, calculated along a surface of constant W_i, and evaluated at the equilibrium point. Therefore in equilibrium, for each principal, the slopes of the payment function equal the slopes of that principal's indifference surface. In other words, the principal's marginal payment for each component of action equals his valuation of that component. This property is similar to the truthful revelation of people's valuation of public goods, externalities etc. under suitably designed mechanisms; hence the name. It holds for any interior equilibrium supported by differentiable payment schedules.

Proof of Proposition 1: Given differentiability and the fact that the equilibrium is interior, we have the first-order conditions for the maximization in (b) of Theorem 1:

$$\frac{\partial G[\mathbf{a}^\circ, \mathbf{C}^\circ(\mathbf{a}^\circ)]}{\partial a_k} + \sum_{j \in L} \frac{\partial G[\mathbf{a}^\circ, \mathbf{C}^\circ(\mathbf{a}^\circ)]}{\partial c_j} \frac{\partial C_j^\circ(\mathbf{a}^\circ)}{\partial a_k} = 0 \qquad (2.4)$$

for all components k of the action vector. Similarly, the first-order conditions for the maximization in (c) of Theorem 1 are

$$\frac{\partial W_i[\mathbf{a}^\circ, C_i^\circ(\mathbf{a}^\circ)]}{\partial a_k} + \lambda \left[\frac{\partial G[\mathbf{a}^\circ, \mathbf{C}^\circ(\mathbf{a}^\circ)]}{\partial a_k} + \sum_{j \in L, \, j \neq i} \frac{\partial G[\mathbf{a}^\circ, \mathbf{C}^\circ(\mathbf{a}^\circ)]}{\partial c_j} \frac{\partial C_j^\circ(\mathbf{a}^\circ)}{\partial a_k} \right] = 0 \qquad (2.5)$$

for all i, k, and

$$\frac{\partial W_i[\mathbf{a}^\circ, C_i^\circ(\mathbf{a}^\circ)]}{\partial c_i} + \lambda \left[\frac{\partial G[\mathbf{a}^\circ, \mathbf{C}^\circ(\mathbf{a}^\circ)]}{\partial c_i} \right] = 0 \qquad (2.6)$$

for all i, where λ is a non-negative Lagrange multiplier for the non-negativity constraint in condition (c). Combining (2.4) and (2.5) yields

$$\frac{\partial W_i[\mathbf{a}^\circ, C_i^\circ(\mathbf{a}^\circ)]}{\partial a_k} = \lambda \; \frac{\partial G[\mathbf{a}^\circ, \mathbf{C}^\circ(\mathbf{a}^\circ)]}{\partial c_i} \; \frac{\partial C_i^\circ(\mathbf{a}^\circ)}{\partial a_k} \; . \tag{2.7}$$

Combining this with (2.6) and stacking up the components into gradient vectors, we get the local truthfulness condition (2.3). □

Next we consider a stronger property of payment functions, namely *global truthfulness*. A globally truthful payment function for principal i rewards the agent for every change in the action exactly the amount of change in the principal's welfare, provided that the payment both before and after the change is strictly positive. In other words, the shape of the payment schedule mirrors the shape of the principal's indifference surface not only near the equilibrium point (as with local truthfulness) but everywhere the payments are positive. Then the principal gets the same utility for all actions $\mathbf{a}$ that induce positive payments $C_i(\mathbf{a}) > 0$; the payment is just the compensating variation. We show that the common agency game has an equilibrium in which all the principals follow globally truthful strategies, and that such an equilibrium is Pareto efficient. We call such an equilibrium a *truthful equilibrium.*

Focus on truthful equilibria may seem restrictive, but can be justified in several different ways. First, note that we do not restrict the space of feasible payment functions to truthful ones at the outset; in a truthful equilibrium each principal's truthful strategy is a best response to his rivals even when the space of feasible payment functions is the larger one of Assumption 1. Thus we have an equilibrium in the full sense, where the strategies happen to be truthful. Second, and more strongly, for any set of feasible strategies of the $L-1$ principals other than i, the set of best response strategies for principal i contains a truthful payment function. Thus each principal bears essentially no cost from playing a truthful strategy, no matter what he expects from the other players. Then the result that an equilibrium in truthful strategies implements a Pareto efficient action may make such strategies focal for the group of principals. Finally, since the setting has no incomplete information, the players have "nothing to hide" and truthful strategies provide a simple device to achieve efficiency without any player conceding his right to grab as much as he can for himself.

We now proceed to formalize the idea and the results.

Definition 3: A payment function $C_i^T(\mathbf{a}, W_i^*)$ for principal i is *globally truthful* relative to the constant W_i^* if

$$C_i^T(\mathbf{a}, W_i^*) \equiv \min[\bar{C}_i(\mathbf{a}), \max[0, \varphi_i(\mathbf{a}, W_i^*)]] \ \ \text{for all } \mathbf{a} \in \mathcal{A}, \tag{2.8}$$

where φ_i is implicitly defined by $W_i[\mathbf{a}, \varphi_i(\mathbf{a}, W_i^*)] = W_i^*$ for all $\mathbf{a} \in \mathcal{A}$, and $\bar{C}_i(\mathbf{a}) = \sup\{C_i(\mathbf{a}) \mid C_i \in \mathcal{C}_i\}$ for all $\mathbf{a} \in \mathcal{A}$.

Explanation: The definition of φ_i is the basic concept of the compensating variation. Equation (2.8) merely serves to ensure that the truthful payment function also satisfies the upper and lower bounds on feasible payments. Note that a competition in truthful strategies boils down to non-cooperative choices of the constants $\{W_j^*\}_{j \in L}$, which determine the equilibrium payoffs of the principals.

Proposition 2: The best response set of principal i to payment functions $\{C_j^\circ(\mathbf{a})\}_{j \in L, j \neq i}$ of the other principals contains a globally truthful payment function.

Explanation: The proposition can be illustrated in the aforementioned Figure 1. The principals other than i induce in the agent the indifference curve $G_i G_i$ with their payment offers. These offers might be truthful or not. In any event, the best response set for principal i includes all payment functions that induce the action and contribution associated with point A. A truthful strategy in this set is the payment function that coincides with the horizontal axis from the origin until its intersection with $W_i W_i$, and that coincides with $W_i W_i$ thereafter.

Proof of Proposition 2: Let C_i° be some best response of principal i to payment functions $\{C_j^\circ(\mathbf{a})\}_{j \in L, j \neq i}$ of the other principals, let $\mathbf{a}^\circ$ be the agent's best response to the whole set of payment functions $\{C_j^\circ(\mathbf{a})\}_{j \in L}$, and let $W_i[\mathbf{a}^\circ, C_i^\circ(\mathbf{a}^\circ)] = W_i^\circ$ be the resulting utility levels of the principals. Define the globally truthful payment function $C_i^T(\mathbf{a}, W_i^\circ)$ relative to W_i°. We claim that it is also a best response to the same given payment functions of all the others.

To see this, let $\mathbf{a}'$ denote the agent's choice of action

$$\mathbf{a}' = \arg\max_{\mathbf{a} \in \mathcal{A}} G[\mathbf{a}, (\{C_j^\circ(\mathbf{a})\}_{j \neq i}, C_i^T(\mathbf{a}, W_i^\circ))].$$

If $\mathbf{a}' = \mathbf{a}^\circ$, then the truthful strategy trivially yields to principal i the same utility level W_i° as does the strategy C_i° which is a best response, so the truthful strategy must itself be a

best response. If $\mathbf{a}' \neq \mathbf{a}^\circ$, this must be because, other things equal, the truthful strategy elicits a larger payment. For if $C_i^\circ(\mathbf{a}') \geq C_i^T(\mathbf{a}', W_i^\circ)$, we have

$$
\begin{aligned}
G[\mathbf{a}', \mathbf{C}^\circ(\mathbf{a}')] \;\; &\geq \;\; G[\mathbf{a}', (\{C_j^\circ(\mathbf{a}')\}_{j \neq i}, C_i^T(\mathbf{a}', W_i^\circ))] \\
&> \;\; G[\mathbf{a}^\circ, (\{C_j^\circ(\mathbf{a}^\circ)\}_{j \neq i}, C_i^T(\mathbf{a}^\circ, W_i^\circ))] \\
&= \;\; G[\mathbf{a}^\circ, \mathbf{C}^\circ(\mathbf{a}^\circ)].
\end{aligned}
$$

The first step follows from the fact that G is increasing in payments, the second because $\mathbf{a}'$ is a maximizer of G while $\mathbf{a}^\circ$ is not, and the third because $C_i^T(\mathbf{a}^\circ, W_i^\circ) = C_i^\circ(\mathbf{a}^\circ)$. Then $\mathbf{a}^\circ$ would not have been the agent's best choice in equilibrium.

We have shown $C_i^T(\mathbf{a}', W_i^\circ) > C_i^\circ(\mathbf{a}') \geq 0$, and there remain two logical cases to consider: $C_i^T(\mathbf{a}', W_i^\circ)$ can be equal to $\varphi_i(\mathbf{a}', W_i^\circ)$, or at the upper limit $\bar{C}_i(\mathbf{a}')$. In the first of these, principal i gets utility W_i°, which is the same as he gets with C_i°, so C_i^T is also a best response. In the second, principal i would have been willing to pay even more while still getting at least W_i°; but then C_i° could not have been the best response and this case cannot arise. $\square$

Definition 4: A *truthful equilibrium* is an equilibrium in which all payment functions are globally truthful relative to the equilibrium welfare levels.

Proposition 3: Let $(\{C_i^T\}_{i \in L}, \mathbf{a}^\circ)$ be a truthful equilibrium in which W_i° is the equilibrium utility level of principal i, for all $i \in L$. Then $(\{W_i^\circ\}_{i \in L}, \mathbf{a}^\circ)$ is characterized by:

(a) $\mathbf{a}^\circ = \arg\max_{\mathbf{a} \in \mathcal{A}} G[\mathbf{a}, \{C_i^T(\mathbf{a}, W_i^\circ)\}_{i \in L}];$

(b) for every $i \in L$,

$$
G[\mathbf{a}^\circ, \{C_i^T(\mathbf{a}^\circ, W_i^\circ)\}_{i \in L}] = \max_{\mathbf{a} \in \mathcal{A}} G[\mathbf{a}, (\{C_j^T(\mathbf{a}, W_i^\circ)\}_{j \neq i}, 0)].
$$

Explanation: This is just a restatement of Corollary 1 for the case of truthful equilibria, and the explanation given above applies. The added advantage is in actual use. If we tried to use Corollary 1 to determine equilibria, we would have to solve the conditions simultaneously for all the payment *functions*, which is a complicated fixed point problem and has a large multiplicity of solutions. The corresponding set of equations in Proposition 3 involve the equilibrium utility *numbers*; therefore they constitute a simpler simultaneous

equation problem with solutions that are in general locally determinate, and in applications often unique. We will consider one such application in the next section.

Proof of Proposition 3: This follows directly from condition (b) in the definition of an equilibrium, from Corollary 1, and from the definition of globally truthful payment functions. □

Now we establish that an equilibrium in globally truthful strategies implements an efficient action.

Theorem 2: Let a policy vector $\mathbf{a}^\circ$ and a vector of payment functions $\mathbf{C}^\circ$ that are globally truthful with respect to the utility levels $W_i^\circ = W_i(\mathbf{a}^\circ, C_i^\circ(\mathbf{a}^\circ))$ constitute a truthful equilibrium. Then there do not exist an action $\mathbf{a}^*$ and a payment vector $\mathbf{c}^*$ such that

(i) feasibility:
$$\mathbf{a}^* \in \mathcal{A}; \qquad 0 \leq c_i^* \leq \bar{C}_i(\mathbf{a}*) \text{ for all } i \in L \,;$$

(ii) Pareto superiority:
$$G(\mathbf{a}^*, \mathbf{c}^*) \geq G[\mathbf{a}^\circ, \mathbf{C}^\circ(\mathbf{a}^\circ)] \,,$$

$$W_i(\mathbf{a}^*, c_i^*) \geq W_i[\mathbf{a}^\circ, C_i^\circ(\mathbf{a}^\circ)] \qquad \text{for all } i \in L \,,$$

with at least one strict inequality.

Explanation: The efficiency of truthful equilibria extends a similar result proved by Bernheim and Whinston (1986, Theorem 2) for the case of transferable utility. We can provide a familiar interpretation by invoking only local truthfulness. This is easiest to see in the case where the agent's preferences depend only on the sum of the total payments, $c = \Sigma_{j \in L} \, c_j$. Then (2.4), (2.6), and (2.7) combine to yield

$$\frac{\partial G/\partial a_k}{\partial G/\partial c} = \sum_{j \in L} \frac{\partial W_j/\partial a_k}{\partial W_j/\partial c_j} \,,$$

or

$$\left(\frac{dc}{da_k} \right)_{G \text{ constant}} = \sum_{j \in L} \left(\frac{dc}{da_k} \right)_{W_j \text{ constant}} \,.$$

This says that the marginal payment the agent requires for supplying an additional unit of action k (the marginal cost) equals the sum of all the principals' marginal willingness to pay for this unit. That is just the Samuelson optimality condition for the supply of a public good,

which is the appropriate interpretation here, because the action is a public good affecting all players. The agent's maximization ensures equality between the agent's marginal cost and the sum of the slopes of the payment schedules, whereas the truthfulness property ensures equality between the slopes of the schedules and the principals' marginal utilities. We leave it to the readers to develop a similar interpretation in the more general case.

Proof of Theorem 2: Suppose there do exist $\mathbf{a}^*$ and $\mathbf{c}^*$ as stipulated in the statement. For all i, we have

$$W_i(\mathbf{a}^*, c_i^*) \geq W_i[\mathbf{a}^\circ, C_i^\circ(\mathbf{a}^\circ)] = W_i[\mathbf{a}^*, \varphi_i(\mathbf{a}^*, W_i^\circ)],$$

by the definition of φ, so $c_i^* \leq \varphi_i(\mathbf{a}^*, W_i^\circ)$. This also ensures $\varphi_i(\mathbf{a}^*, W_i^\circ) \geq 0$, and we have assumed $c_i^* \leq \bar{C}_i(\mathbf{a}^*)$, so

$$c_i^* \leq \min\left[\,\bar{C}_i(\mathbf{a}^*), \max[0, \varphi_i(\mathbf{a}^*, W_i^\circ)]\,\right] = C_i^\circ(\mathbf{a}^*)$$

using the definition of a truthful schedule.

Since G is increasing in payments, this shows $G[\mathbf{a}^*, \mathbf{C}^\circ(\mathbf{a}^*)] \geq G(\mathbf{a}^*, \mathbf{c}^*)$. Also, since $\mathbf{a}^\circ$ is the agent's best response to $\mathbf{C}^\circ$, we have $G[\mathbf{a}^\circ, \mathbf{C}^\circ(\mathbf{a}^\circ)] \geq G[\mathbf{a}^*, \mathbf{C}^\circ(\mathbf{a}^*)]$. If the agent's utility inequality is strict in (ii) of the statement of the theorem, we already have a contradictory chain of inequalities. So consider the case where

$$G(\mathbf{a}^*, \mathbf{c}^*) = G[\mathbf{a}^\circ, \mathbf{C}^\circ(\mathbf{a}^\circ)] = G[\mathbf{a}^*, \mathbf{C}^\circ(\mathbf{a}^*)]\,.$$

We prove that this leads to another contradiction if the utility inequality in (ii) is strict for any one principal i.

Consider the constraint (2.2) in Condition (c) of Theorem 1. It is satisfied by $[\mathbf{a}^\circ, \mathbf{C}^\circ(\mathbf{a}^\circ)]$ and therefore by $(\mathbf{a}^*, \mathbf{c}^*)$ in the case we are now considering. But for all principals $j \neq i$, we have already established $c_j^* \leq C_j^\circ(\mathbf{a}^*)$, and G is increasing in every component of the payment vector. Therefore the constraint is also satisfied by $[\mathbf{a}^*, (\{C_j^\circ(\mathbf{a}^*)\}_{j \neq i}, c_i^*)]$. Thus $(\mathbf{a}^*, c_i^*)$ is feasible in the maximization problem in Condition (c) of Theorem 1. But we are supposing that it gives strictly more utility than $W_i[\mathbf{a}^\circ, C_i^\circ(\mathbf{a}^\circ)]$. This contradicts Condition (c) of Theorem 1, which $\circ$ being an equilibrium must satisfy. $\square$

Quasi-Linear Preferences

The above equilibrium can be pinned down in greater detail when all the players' preferences are linear in the payments. Specifically, the action is independent of the distribution in this case.

Corollary 2: Let the preference functions $(\{W_i\}_{i \in L}, G)$ be of the quasi-linear form

$$W_i(\mathbf{a}, c_i) = w_i(\mathbf{a}) - \omega_i c_i \qquad \text{for all } i \in L,$$

and

$$G(\mathbf{a}, \mathbf{c}) = g(\mathbf{a}) + \gamma \sum_{i \in L} c_i.$$

Consider a truthful equilibrium where the action is $\mathbf{a}^\circ$ and all payments are in the interior: $0 < C_i^\circ(\mathbf{a}^\circ) < \bar{C}_i(\mathbf{a}^\circ)$. Then

$$\mathbf{a}^\circ = \arg \max_{\mathbf{a} \in \mathcal{A}} \ g(\mathbf{a})/\gamma + \sum_{i \in L} w_i(\mathbf{a})/\omega_i.$$

Proof of Corollary 2: If not, a better neighboring feasible action, plus a suitable rearrangement of payments, can achieve a Pareto superior outcome, contradicting Theorem 2. □

Explanation: With quasi-linear preferences, the equilibrium action maximizes a weighted sum of gross welfare levels of the principals and the agent. This result has been useful in applications to political economy, such as in Grossman and Helpman (1994). There, the agent is a government that sets a vector of tariff policies, while the principals are interest groups representing owners of sector-specific factors of production. The government's objective is assumed to be linear in the aggregate welfare of voters and the total of campaign contributions collected from special interests. The corollary predicts a structure of protection that maximizes a simple weighted sum of the welfare of voters and interest-group members.

Government Policy-Making

As we noted in the introduction, common agency arises frequently in the political processes that generate economic policies. The policy makers often can be viewed as an agent and some or all of their constituents as principals. Principals can "lobby" the policy makers

by promising payments in return for policies, within some prescribed limits on available policies and feasible donations. The payments may take the form of illicit bribes or, more typically, implicit (and therefore legal) offers of campaign support. The Grossman-Helpman application to tariff-setting is but one example of this. In such settings, it may be natural to think of the government as having an objective function with social welfare and the total of contribution receipts as arguments. The government might care about social welfare for ethical reasons, or it may want to provide a high standard of living to enhance its re-election prospects, or to keep the populace sufficiently happy to prevent riots, etc. Contributions likewise might enter the government's objective because they affect its re-election chances, or merely as utility of the governing elites' private consumption. Accordingly, we suppose $G(\mathbf{a}, \mathbf{c}) = g(w, \bar{c})$, where

$$w = W[u_1(\mathbf{a}, c_1), u_2(\mathbf{a}, c_2) \ldots u_n(\mathbf{a}, c_n)]$$

is a Bergson-Samuelson social welfare function which is increasing in the utility levels of the n individuals in the polity, and $\bar{c} = \sum_{i=1}^{n} c_i$ is the aggregate contribution.

Let L be the set of individuals who can lobby the government for special favors. We leave L exogenous – some individuals may have special connections, or some groups of individuals may be able to solve the free rider problem of organized action while others cannot. Then $C_i(\mathbf{a}) \equiv 0$ for $i \notin L$. For $i \in L$, the upper limit on feasible contributions, $\bar{C}_i(\mathbf{a})$, is implicitly defined by $u_i[\mathbf{a}, \bar{C}_i(\mathbf{a})] = \underline{u}_i$, where $\underline{u}_i$ is the lowest or subsistence utility level for individual i.

Theorem 2 has strong implications for the outcome of this lobbying game.

Corollary 3: Let the agent's preferences be given by $G(\mathbf{a}, \mathbf{c}) = g(w, \bar{c})$, where $w = W[u_1(\mathbf{a}, c_1), u_2(\mathbf{a}, c_2), \ldots, u_n(\mathbf{a}, c_n)]$ and $\bar{c} = \sum_{i=1}^{n} c_i$. Let a set $L \subset \{1, 2, \ldots, n\}$ of individuals offer payment schedules $\{C_i(\mathbf{a})\}_{i \in L}$, while $C_i(\mathbf{a}) \equiv 0$ for $i \notin L$. Finally, let a policy vector $\mathbf{a}^\circ$ and a vector of payment functions $\mathbf{C}^\circ$ which are globally truthful with respect to the utility levels $u_i^\circ = u_i[\mathbf{a}^\circ, C_i^\circ(\mathbf{a}^\circ)]$ for $i \in L$ constitute a truthful equilibrium in which $u_i^\circ = u_i(\mathbf{a}^\circ, 0)$ for $i \notin L$. Then there exists no other policy vector $\mathbf{a}'$ such that $u_i(\mathbf{a}', c_i^\circ) \geq u_i^\circ$ for all $i \in \{1, 2, \ldots, n\}$ with strict inequality holding for some i.

Proof of Corollary 3: Suppose, to the contrary, that there exists a vector $\mathbf{a}'$ such that $u_i(\mathbf{a}', c_i^\circ) \geq u_i^\circ$ for all $i \in \{1, 2, \ldots, n\}$, with strict inequality for some i. Since the con-

tributions c_i° were feasible when the action was $\mathbf{a}^\circ$, we know $u_i(\mathbf{a}^\circ, c_i^\circ) \geq \underline{u}_i$. A fortiori, $u_i(\mathbf{a}', c_i^\circ) \geq \underline{u}_i$; thus the contributions c_i° satisfy $c_i^\circ \leq \bar{C}_i(\mathbf{a})$, and are therefore feasible.

Since $w(\cdot)$ is increasing in all of its arguments, the individual utility inequalities imply that the government also gets greater utility using $\mathbf{a}'$:

$$g\{W[u_1(\mathbf{a}', c_1^\circ), u_2(\mathbf{a}', c_2^\circ) \ldots u_n(\mathbf{a}', c_n^\circ)], \mathbf{c}^\circ\} > g[W(u_1^\circ, u_2^\circ, \ldots, u_n^\circ), \mathbf{c}^\circ]\} \, .$$

We have found a feasible and Pareto superior alternative pair of action and contribution vectors, which violates part (ii) of Theorem 2 for a truthful equilibria. $\square$

Explanation: The corollary says that, even under the pressure of lobbying from a set of special interests, a government that has some concern for social welfare will make Pareto efficient choices from the set of feasible policies. With truthful payment schedules, the government has incentive to collect its tributes efficiently. If the government's objective weighs positively the well-being of all members in society, then efficiency for the government and lobbyers translates into Pareto efficiency for the polity as a whole. We will see an application of this in the next section.

To summarize, our general theory preserves the flavor of results in Bernheim and Whinston (1986). In any subgame perfect Nash equilibrium, the agent's action and the payment from principal i maximize the joint welfare of the agent and that principal, given the equilibrium payment functions of the other principals. This is similar to their Lemma 2. For any set of offers by principals $j \neq i$, the best response set for principal i contains a truthful strategy, just as in their Theorem 1. And the truthful equilibrium that results when all principals announce truthful payment functions is Pareto efficient for the group of principals and the agent. By extending their results, we have significantly expanded the domain of their theory.

In the next section we show the usefulness of the general theory in a specific application. We study a positive analog to the normative theory of tax setting à la Diamond and Mirrlees (1971). We extend their analysis to the case of a government that cares not only about aggregate welfare, but also about the contributions it can amass. The endogenous taxes and transfers are those that arise as an equilibrium in a common-agency game where the special interests bid for favored treatment.

3   Application to Tax and Transfer Policies

We closely follow Diamond and Mirrlees (1971), but introduce lobbying for taxes and transfers. Let there be n consumers, labelled $i \in N = \{1, 2, \ldots, n\}$. We will continue to refer to these as "individuals", but in reality most lobbying is undertaken by groups of individuals who have overcome the free rider problem of collective political action. If such special interest groups have access to internal transfer schemes, they can be regarded as aggregated individuals in our model.

Only the subset L of individuals can lobby the government. These may be the players with the largest stake in the policies, or those with personal connections to the politicians. As before, lobbying takes the form of contingent contributions. The lobbyers hope to influence the government's policy choices.

To simplify the exposition we consider a small open economy.[4] Let $\mathbf{p}^w$ denote the exogenous vector of world prices, and $\mathbf{q}$, $\mathbf{p}$ the price vectors faced by the domestic consumers and producers respectively. Then $\mathbf{q} - \mathbf{p}^w$ is the implied vector of consumer tax rates (negative components are subsidies), and $\mathbf{p} - \mathbf{p}^w$ the implied vector of producer subsidy rates (negative components are taxes). The government's tax and subsidy policies are therefore equivalent to choosing $\mathbf{q}$ and $\mathbf{p}$. The government can also make lump-sum transfers or levy lump-sum taxes on any or all individuals; let $\mathbf{t}$, with components t_i, for $i \in N$ denote the vector of such transfers (negative components are taxes). We leave out any other government activities for simplicity.

There are several firms labelled $f \in M = \{1, 2, \ldots m\}$ with profit functions $\psi^f(\mathbf{p})$, and by Hotelling's Lemma, supply functions $\nabla_{\mathbf{p}} \psi^f(\mathbf{p})$. Individual i owns an exogenous share ω_{if} of firm f, and therefore gets profit income

$$\pi^i(\mathbf{p}) = \sum_{f \in M} \omega_{if}\, \psi^f(\mathbf{p})\,.$$

Purely for notational convenience, we define

$$S^i(\mathbf{p}) \equiv \nabla_{\mathbf{p}}\, \pi^i(\mathbf{p}) = \sum_{f \in M}\, \omega_{if}\, \nabla_{\mathbf{p}}\, \psi^f(\mathbf{p})\,,$$

---

[4]All the results concerning efficiency have identical counterparts when some or all commodities are non-tradeable; all that is needed is some additional algebra for the domestic market-clearing conditions.

the supply attributable to the fractional firms owned by individual i. We refer to this for brevity as the supply "from" individual i, although individuals as such do not do any production or supplying. The total supply is

$$S(\mathbf{p}) = \sum_{i \in N} S^i(\mathbf{p}) \,.$$

Let c_i denote the lobbying payment of individual i to the government, for $i \in L$. Set $c_i \equiv 0$ for $i \notin L$. Then individual i's income is

$$I_i \equiv \pi^i(\mathbf{p}) + t_i - c_i \,, \tag{3.1}$$

and his resulting indirect utility is

$$u_i = V^i(\mathbf{q}, I_i) \,.$$

His demands are given by Roy's Identity:

$$D^i(\mathbf{q}, I_i) = -\nabla_{\mathbf{q}} V^i(\mathbf{q}, I_i) / V_I^i(\mathbf{q}, I_i) \,.$$

We should emphasize that the payments made by the lobbies do not enter into the government's tax and transfer budget. This budget reflects the "public" or policy part of the government's activities. The lobbies' payments go into a separate "private" or political kitty. They might be used by the governing party for its re-election campaign, or by a governing dictator for his own consumption. Write $\mathbf{g}$ for the vector of such purchases. Thus there are two budget constraints, one for the policy budget

$$(\mathbf{q} - \mathbf{p}^w)' \sum_{i \in N} D^i(\mathbf{q}, I_i) - (\mathbf{p} - \mathbf{p}^w)' S(\mathbf{p}) - \sum_{i \in N} t_i = 0 \,.$$

and the other for the political budget

$$\mathbf{p}^{\mathbf{w}\prime} \mathbf{g} = \sum_{i \in L} c_i \,.$$

These constitute only one independent equation, because one can be derived from the other using the individuals' and the producers' budget identities and the economy's trade balance condition. We will use the policy budget condition, and aggregate the political purchases

at the world prices so only the sum of the contributions enters the government's objective function.[5]

We can now regard the government as choosing $\mathbf{a} = (\mathbf{q}, \mathbf{p}, \mathbf{t})$ subject to a budget constraint. This puts the problem in the framework of the subsection on Government Policy-Making above. Corollary 3 there immediately gives a strong result concerning the choice of action in a truthful equilibrium of the policy game.

Efficiency

The corollary says that the equilibrium action achieves a Pareto efficient outcome in an auxiliary problem where the lobbies' payments are held fixed at their equilibrium levels. In the auxiliary problem, the government's choice is the standard normative optimal tax and transfer problem, where we know that if lump-sum transfers are available, distorting commodity taxes and subsidies will not be used. Therefore we have shown that the political equilibrium will also preserve $\mathbf{q} = \mathbf{p}$ and use only the lump-sum transfers $\mathbf{t}$ for the two purposes of eliciting contributions from the lobbies and of meeting the government's concern for the welfare of the non-lobbying individuals.

Before the readers form the belief that we have established the efficiency of the political process of tax policy, however, we should warn that the story is not yet complete. It remains to examine the distribution of gains between the lobbies and the government in the political equilibrium; that analysis will cast doubt on the efficient equilibrium as a description of political reality.

We first develop the truthful equilibrium in somewhat greater detail, and elaborate upon the efficiency property. This will serve several purposes. First, it will clarify the mechanism that makes it more desirable for the government to collect its tributes from the lobbies without introducing any distortions and thus strengthen the economic intuition for the efficiency result. It will also set the stage for the analysis of distribution to come. Finally, it will establish a better connection with previous work on lobbying for tariffs (Grossman and Helpman 1994) and for commodity taxes alone (Dixit 1995).

We construct a truthful equilibrium with endogenous policies $(\mathbf{q}^\circ, \mathbf{p}^\circ, \mathbf{t}^\circ)$, utility levels $\{u_i^\circ\}_{i \in N}$, and truthful payment schedules $\{C_i^\circ(\mathbf{q}, \mathbf{p}, t_i, u_i^\circ)\}_{i \in L}$.

---

[5]We have assumed that the government's political purchases do not pay the tax; the alternative of making them subject to tax would complicate the algebra somewhat but would not alter the conclusions.

The compensating variations φ_i relative to this equilibrium are defined by

$$V^i(\mathbf{q}, \pi^i(\mathbf{p}) + t_i - \varphi_i) = u_i^\circ$$

or

$$\varphi_i(\mathbf{q}, \mathbf{p}, t_i, u_i^\circ) = \pi^i(\mathbf{p}) + t_i - E^i(\mathbf{q}, u_i^\circ),$$

where E^i is the expenditure function dual to the indirect utility function. The truthful schedules relative to the equilibrium are

$$C_i^\circ(\mathbf{q}, \mathbf{p}, t_i, u_i^\circ) = \max\left[0, \varphi_i(\mathbf{q}, \mathbf{p}, t_i, u_i^\circ)\right].$$

Suppose that the equilibrium contribution levels

$$c_i^\circ = C_i^\circ(\mathbf{q}^\circ, \mathbf{p}^\circ, t_i^\circ, u_i^\circ)$$

are zero for a set P of individuals i (the "passive" lobbies), and are positive (and therefore equal to the compensating variations) for a set $A = L - P$ (the "active" lobbies). Of course P and A are determined endogenously in the equilibrium. We characterize the equilibrium using Proposition 3 of the previous section. Condition (a) of that proposition determines the government's policy choice, and condition (b) determines the contributions of the lobbies. We examine these in turn, under the headings of efficiency and distribution. While efficiency has already been proved, we now show more explicitly how an inefficient set of policies can be improved upon.

Therefore suppose that domestic prices in equilibrium differ from world prices.[6] We prove that, given the equilibrium payment schedules of the individuals in L, the government can do at least as well (and generally better) with a non-distorting policy $(\mathbf{p}^w, \mathbf{p}^w, \mathbf{t}^1)$ which keeps all domestic prices equal to the corresponding world prices, and chooses transfers suitably. This violates Condition (a) of Proposition 3.

We define the alternative transfers by

$$V^i(\mathbf{p}^w, \pi^i(\mathbf{p}^w) + t_i^1 - c_i^\circ) = u_i^\circ$$

---

for all $i \in N$. Here we have $c_i^\circ \equiv 0$ for the individuals $i \in N - L$ who are simply not allowed to lobby by the rules of the game; for $i \in L$ the c_i° are as stated above.

Given this alternative policy of the government, all individuals have the same compensating variations as before, and therefore all those in L will contribute the same amounts as before, zero for those in P whose compensating variations are non-positive, and given by the compensating variations $c_i^1 = c_i^\circ > 0$, for the $i \in A$. All get the same utilities as before, so the government's objective function takes the same value as before. It remains to show that the policy satisfies the government's budget constraint.

Rewrite the definition of the transfers as

$$ t_i^1 = E^i(\mathbf{p}^w, u_i^\circ) - \pi^i(\mathbf{p}^w) + c_i^\circ \,. $$

Now note that, by convexity of the profit functions,

$$
\begin{aligned}
\pi^i(\mathbf{p}^w) &\geq \pi^i(\mathbf{p}^\circ) + (\mathbf{p}^w - \mathbf{p}^\circ)' S^i(\mathbf{p}^\circ) \\
&= \pi^i(\mathbf{p}^\circ) - (\mathbf{p}^\circ - \mathbf{p}^w)' S^i(\mathbf{p}^\circ) \,.
\end{aligned}
$$

Similarly, by concavity of the expenditure functions,

$$
\begin{aligned}
E^i(\mathbf{p}^w, u_i^\circ) &\leq E^i(\mathbf{q}^\circ, u_i^\circ) + (\mathbf{p}^w - \mathbf{q}^\circ)' \bar{D}^i(\mathbf{q}^\circ, u_i^\circ) \\
&= E^i(\mathbf{q}^\circ, u_i^\circ) - (\mathbf{q}^\circ - \mathbf{p}^w)' D^i(\mathbf{q}^\circ, I_i^\circ)
\end{aligned}
$$

because the levels of Hicksian compensated demands $\bar{D}^i$ and the uncompensated demands D^i are equal at the original equilibrium $\circ$. If i is either in the non-lobbying set $(N - L)$ or in the passive lobbying set P, we have $c_i^\circ = 0$ so

$$ E^i(\mathbf{q}^\circ, u_i^\circ) = I_i = \pi^i(\mathbf{p}^\circ) + t_i^\circ \,. $$

Combining all these equations and inequalities, we get

$$ t_i^1 \leq t_i^\circ - (\mathbf{q}^\circ - \mathbf{p}^w)' D^i(\mathbf{q}^\circ, I_i^\circ) + (\mathbf{p}^\circ - \mathbf{p}^w)' S^i(\mathbf{p}^\circ) \,. $$

If i is in the active lobbying set A, then c_i° is given by the compensating variation above. Again we combine the relevant equations, and the result is the same inequality for t_i^1 as the one above.

Finally, adding these inequalities for all $i \in N$,

$$\sum_{i \in N} t_i^1 \leq \sum_{i \in N} t_i^\circ - (\mathbf{q}^\circ - \mathbf{p}^w)' \sum_{i \in N} D^i(\mathbf{q}^\circ, I_i^\circ) + (\mathbf{p}^\circ - \mathbf{p}^w)' \sum_{i \in N} S^i(\mathbf{p}^\circ) \leq 0 \,,$$

using the budget constraint which is satisfied by the equilibrium $\circ$. This proves the feasibility of the alternative – no-distortions, transfers only – policy.

If $\mathbf{q}^\circ$ and $\mathbf{p}^\circ$ are not both equal to $\mathbf{p}^w$, and there is some substitution possibility in either consumption or production for at least some i, then at least one of the inequalities on the profit and expenditure functions will be strict. Then the alternative policy can be turned into a strictly superior response for the government. It can increase t_i^1 for any $i \in N - L$ or $i \in P$ slightly, which increases u_i^1, or increase t_i^1 for any $i \in A$, which increases c_i^1, in each case leaving all the other arguments of the government's objective function unchanged.

The argument is very familiar from standard analyses of gains from trade, for example Dixit and Norman (1980, chapter 3). Removal of a distortion gives consumers the opportunity to achieve the same utility at a lower expenditure, and producers the opportunity to increase their profits. The government can then achieve the same outcome while handing out smaller transfers, or more pertinently, can achieve a better outcome by increasing transfers. In the normative analysis, the government cared only about the resulting increase in individuals' utilities; in the political equilibrium it also cares about the increase in the contributions it gets from the lobbies.

Distribution

Condition (b) of Proposition 3 helps us calculate the individuals' utility levels u° and the government's receipts from the lobbies. The condition says that the government's utility in equilibrium should equal what it would get by responding optimally to the equilibrium payment schedules of all the lobbies except any one, when that one pays nothing. The set of equations this generates are to be solved simultaneously. That is in general intractable, and we consider a somewhat restricted problem where the social welfare function and the government's objective are both additive:

$$w = W(u_1, u_2, \ldots u_n) = u_1 + u_2 + \ldots u_n \,,$$

and

$$G(w, \bar{c}) = w + \theta \, \bar{c} \,.$$

24

As is usual in this situation, we assume that the individuals' utility functions $V^i(\mathbf{q}, I_i)$ are strictly concave in I_i, and the marginal utilities $V^i_I(\mathbf{q}, I_i)$ go to ∞ as $I_i \to 0$, thus yielding an interior solution to the government's maximization problem.

Using the efficiency result of the previous subsection, we set $\mathbf{p} = \mathbf{q} = \mathbf{p}^w$, and omit these arguments from the various functions. Define $\pi_i = \pi^i(\mathbf{p}^w)$, and think of them as the individuals' endowments. Then the non-lobbyers' incomes are $\pi_i + t_i$, the lobbies' truthful contribution schedules are $c_i = \pi_i + t_i - E^i(u^\circ_i)$, and the government's choice of $\mathbf{t}$ maximizes

$$g = \sum_{i \in L} u^\circ_i + \sum_{i \notin L} V^i(\pi_i + t_i) + \theta \sum_{i \in L} [\pi_i + t_i - E^i(u^\circ_i)] \qquad (3.2)$$

subject to $\Sigma_{i \in N} t_i = 0$.

The first-order conditions for $i \in L$ are just $\theta = \lambda$, which defines the Lagrange multiplier on the budget constraint. Then the conditions for $i \notin L$ become

$$V^i_I(\pi + t_i) = \theta \,,$$

or

$$\pi_i + t_i = H^i(\theta) \,, \qquad (3.3)$$

where H^i is the inverse of individual i's marginal utility function. This gives the consumption level of individual i when this individual is not lobbying but someone else is.

Adding (3.3) across individuals and using the government's budget constraint, we have

$$\sum_{i \in L} (\pi_i + t_i) = \Pi - \sum_{i \notin L} H^i(\theta) \,,$$

where $\Pi \equiv \Sigma_{i \in N} \pi_i$. Using this in the truthful payment schedules of the lobbies, the government's optimized objective becomes

$$g = \sum_{i \in L} u^\circ_i + \sum_{i \notin L} V^i(H^i(\theta)) + \theta \left[\Pi - \sum_{i \notin L} H^i(\theta) - \sum_{i \in L} E^i(u^\circ_i) \right] . \qquad (3.4)$$

Now we can use Condition (b) of Proposition 3. Conduct the following thought experiment. Transfer any one individual j from the lobbying set L to the non-lobbying set $N - L$. Let all the other lobbying individuals adhere to their payment schedules in the original truthful equilibrium. What is the government's welfare? The answer is already known; we just use the above formula (3.4), shifting j from L to $N - L$. Then Condition (c) tells us to

25

equate the value of the government's objective that results from this thought experiment to the value in the original problem. The additive separability makes this simple:

$$u_j^\circ - \theta\, E^j(u_j^\circ) = V^j[H^j(\theta)] - \theta\, H^j(\theta)\,.$$

This is obviously satisfied by $u_j^\circ = V^j[H^j(\theta)]$. But the left hand side is a concave function of u_j and achieves its maximum when $1 = \theta\, dE^j/du_j$, that is when the marginal utility of consumption is θ, or when consumption is $H^j(\theta)$, or $u_j^\circ = V^j[H^i(\theta)]$. Therefore the solution is unique.

The term $V^j[H^i(\theta)]$ describes the utility from the consumption $H^j(\theta)$ that individual j would get if he were genuinely a non-lobbyer and some other individuals were active in lobbying. In the equilibrium, $u_j^\circ = V^j[H^i(\theta)]$; i.e., the lobbies fare no better than they would if they were politically inactive while some others lobbied! This is in fact the worst outcome for individual j, because he can always achieve the same result by *unilaterally* renouncing any lobbying. However, such unilateral renunciation by all lobbies would not be an equilibrium, because starting from such a position each one would want to lobby; that is just the essence of a prisoners' dilemma.

The intuition is that the lobbies are competing for lump-sum transfers, which are available in a constant total sum. The government can implicitly but credibly threaten each lobby that it will cut it out and deal with all the other lobbies, and thereby extract all the surplus that the lobby potentially stands to gain. In other words, the lobbies are in a zero-sum game which turns into a prisoners' dilemma. They would all be better off if they could all commit not to lobby at all, when the government would get no payments and just maximize social welfare. It is the existence of other lobbies that enables the government to push each lobby to an even lower level of welfare.

Note that for our thought experiment to work, there must be at least two lobbies in the initial equilibrium. If there is just one lobby, without it the government can do no better than to maximize social welfare, and in equilibrium the lobby can capture all the surplus inherent in the relationship between the two. But as soon as there are two or more lobbies, each loses all the power and the government gets all the surplus in the form of contributions.

For more general non-linear objective functions, the result can be different in detail. There will still be a prisoners' dilemma in the lobbies competition for transfers, although it may be less severe than in our example where it reduced each lobbyer to the worst possible

outcome. If the policy instruments create distortions, the government given its concern for social welfare is disinclined to use them to extremes. This serves to moderate the lobbies' competition, and allow them to get some gains, as was the case with tariffs in Grossman and Helpman (1994), or production subsidies in Dixit (1995).

Our finding casts the efficiency result of the previous subsection in a very different light. Even though the economy as a whole may exhibit greater efficiency when the lobbies have access to, and therefore obtain in equilibrium, non-distorting lump-sum instruments, the lobbies fare poorly as a result. If they can look ahead, and write ex ante a constitution for lobbying so that they become committed to those rules of the game, they will unanimously agree to a rule which prohibits lobbying for direct transfers, and instead restricts lobbying to distorting policies. Far from justifying an efficient outcome, the result suggests another reason why we might expect inefficient policies to emerge in a political equilibrium.[7]

4   Concluding Comments

In this paper we extended the theory of common agency under complete information to handle situations where preferences are not quasi-linear and utility is non-transferable. We proved that an equilibrium in truthful schedules achieves an outcome that is efficient for the agent and all the principals taken together. The principals' utilities in equilibrium are implicitly determined by a set of simultaneous equations.

We applied this theory to construct a positive model of public finance, where the government chooses commodity taxes and lump-sum transfers in response to lobbying contributions from a subset of individuals. The efficiency result implies that even in this political equilibrium the government uses its available instruments in a Pareto optimal fashion; in particular, it does not use distortionary taxes or subsidies if non-distorting individualized lump-sum transfers are available. However, we found that the lobbies' competition for lump-sum transfers turns into a mutually harmful prisoners' dilemma. If there is an earlier constitution-writing stage of the political game, all lobbies will agree at that point to prohibit the use of such transfers.

---

[7] Hammond (1979) argues that individualized lump-sum transfers are infeasible for informational reasons. See also Coate and Morris (1995) for an informational reason and Dixit and Londregan (1995) for a commitment reason why the political process uses inefficient instruments.

The model promises many fruitful extensions and other applications. In the context of public economics, a positive analysis of policies in an economy with externalities or other distortions seems particularly important. The efficiency result suggests that the political process will use the appropriate Pigovian taxes or subsidies to internalize externalities. But such taxes and subsidies have their own distributional implications. If lump-sum transfers are not available because the constitution of lobbying has ruled them out, or for any of the reasons discussed in Footnote 7 above, then the politics of lobbying for or against Pigovian taxation becomes more intricate and interesting.

Finally, we have specified exogenously the set of individuals who were allowed to lobby. An endogenous theory of the formation of lobbies remains an important part of future research. We believe that our theory will form a useful component of such more general analyses.

References

Bernheim, B. Douglas and Michael Whinston. 1986. "Menu Auctions, Resource Allocation, and Economic Influence," *Quarterly Journal of Economics*, vol. 101, no. 1, February, pp. 1–31.

Coate, Stephen and Stephen Morris. 1995. "On the Form of Transfers to Special Interests." *Journal of Political Economy*, vol. 103, no. 6, December, pp. 1210–1235.

Diamond, Peter A. and James A. Mirrlees. 1971. "Optimal Taxation and Public Production," *American Economic Review*, vol. 61, no. 1, March, pp. 8–27 and no. 4, September, pp. 261–278.

Dixit, Avinash K. 1995. "Special-Interest Lobbying and Endogenous Commodity Taxation." Working paper, Princeton University.

Dixit, Avinash K. and Victor D. Norman. 1980. *Theory of International Trade.* Cambridge, UK: Cambridge University Press.

Dixit, Avinash K. and John B. Londregan. 1995. "Redistributive Politics and Economic Efficiency." *American Political Science Review,* vol. 89, no. 4, December, pp. 856–866.

Grossman, Gene M. and Elhanan Helpman. 1994. "Protection for Sale." *American Economic Review*, vol. 84, no. 4, September, pp. 833–850.

Hammond, Peter J. 1979. "Straightforward Individual Incentive Compatibility in Large Economies." *Review of Economic Studies*, vol. 47, no. 2, April, pp. 263–282.