This is a draft; email me with comments, typos, clarifications, etc.

The Appropriation Principle

The questions below reference the *Appropriation Principle*, which does not seem to appear anywhere in the class notes; however, it is a fairly clear reference to the concepts presented in Ostroy's *Lecture Notes on Incentive Compatibility and Efficiency.* To make it plain what is being referenced, a succinct intuitive definition follows.

Definition

If an allocation mechanism obtains an efficient outcome, a sufficient condition for incentive compatibility of agent reports is that agent surplus is precisely social surplus, plus or minus some transfer which is independent of the agent's report; this is the *appropriation principle.*

A familiar example of the appropriation principle is the VCG mechanism: in VCG, agent utility is given by

$$u_i(\theta_i) = S(\theta) - S_{-i}(\theta_{-i})$$

Note that the term $S_{-i}(\theta_{-i})$ was given by Riley as $S(\alpha_i, \theta_{-i})$, where α_i was the lower bound of the support of agent i's ex ante type distribution; however this definition meshes quite nicely with Ostroy's notion of, "surplus when the agent is not a member of the economy." In either case — restricting our attention to "interesting" problems where there is some potential of the good being produced — agents are rewarded precisely with their contribution to social surplus; we see that the agent's payoff is the social value $S(\theta)$ less some quantity which is not dependent on the agent's reported type, $S_{-i}(\theta_{-i})$.

Traditionally, we apply this concept to solve for the transfer payments which must be issued in order to make a mechanism incentive compatible. This is the approach followed in many Riley questions and in the Ostroy questions below. We generally express value as $u_i(\theta_i) = v_i(\theta) + m_i(\theta)$; rearranging, this gives an explicit form for money payments of $m_i(\theta) = S(\theta) - v_i(\theta) - S_{-i}(\theta_{-i})$.

Of course, there is a difference between Riley's notation and Ostroy's notation. The best advice here is not to get too caught up in the notation; if you can remember the concept of extracting precisely your individual contribution to social surplus — or even the more general form of tracking your surplus with social surplus — it is generally not difficult to slip extant notation into the framework.

Will a monopolist tell the truth?

Suppose a known linear demand $p = A - Bz$ and a single seller with cost function $c(z) = \sigma z$, where marginal cost $\sigma \geq 0$ is known only to the seller (here, it is convenient to let $z \geq 0$ for the seller as well as the buyers). The allocation of z depends on reported [marginal] cost C. Assume $\sigma, C \in [0, A]$.

Consider two possible mechanisms:

[PTE] The allocation $z(C)$ is the PTE quantity and the seller's revenue is $z(C)$ times the PTE price.

[MON] The allocation $z(C)$ is the [simple] monopoly quantity and the seller's revenue is $z(C)$ times the monopoly price.

(a) Construct $\pi_{\text{PTE}}(\sigma, C)$, the profits to the seller under [PTE] with costs σ who reports cost C.

Solution: in price-taking equilibrium, price equals marginal cost. Given our current setup, we can make a further statement that if market price does not equal the [constant] marginal cost C, the firm faces an incentive to produce an infinite amount or to produce nothing (this is no different from differential arguments regarding nonlinear cost functions, but the intuition is much cleaner here). Since the seller's true cost σ is unknown, price-taking equilibrium will be computed according to the stated marginal cost C.

At $p = C$, market demand solves $C = A - Bz$; hence we have $z = \frac{A-C}{B}$. With the seller's profits given by market price less cost times the quantity supplied, we obtain

$$\pi_{\text{PTE}}(\sigma, C) = \left(\frac{A - C}{B}\right)(C - \sigma)$$

(b) Define the incentive compatibility condition for π_{PTE}. For what values, if any, is π_{PTE} incentive compatible?

Solution: for incentive compatibility, we need that the payoff from truthful reporting is at least as large as the payoff from any feasible misreport. Formally, this is stated as

$$\forall C, \quad \pi_{\text{PTE}}(\sigma, \sigma) \geq \pi_{\text{PTE}}(\sigma, C)$$

From part (a), we can see that $\pi_{\text{PTE}}(\sigma, \sigma) = 0$; that is, by truthfully reporting his type (and given the linear cost function) the seller obtains zero profits from a price-taking equilibrium. Then incentive compatibility requires

$$\pi_{\text{MON}}(\sigma, C) \leq 0$$

for all feasible reports C.

Notice that profits are defined by a downward-facing parabola, $\pi_{\text{PTE}}(\sigma, C) = -\frac{1}{B}(C - A)(C - \sigma)$. This parabola has intercepts $C = A$ and $C = \sigma$; thus when $C \in (\sigma, A)$ (by the assumption that $\sigma \in [0, A]$) the firm can obtain positive profits from the mechanism [PTE][1].

It follows that as long as the interval (σ, A) has positive measure the mechanism [PTE] is not incentive compatible; as a corollary, when $\sigma = A$ the mechanism is incentive compatible, due to the fact that optimum profits here — from *any* report — are 0.

(c) Construct $\pi_{\text{MON}}(\sigma, C)$, the profits to the seller under [MON] with costs σ who reports cost C.

Solution: if the seller faces cost C, the monopoly outcome solves

$$\max_z zp(z) - Cz = \max_z z(A - Bz) - Cz$$

First-order conditions give us

$$z = \frac{A - C}{2B}$$

In this case, the clearing price is

$$p = \frac{A + C}{2}$$

It's worth stepping out of this particular subquestion for a moment and remembering what's going on here: the seller *is not* directly selecting the quantity he will produce; he is supplying a report of his marginal cost into a mechanism which then requires him to produce the monopoly level of output. Thus

[1] In section, we made this argument by solving for the optimum C and then determining the sign. This gives us the ability to explicitly calculate the optimal profits of the firm in [PTE] (it turns out they are monopoly profits; this is fairly intuitive); however since this isn't particularly useful and more algebraic result is fair.

the strategic optimization here is *not* over quantity (at least, not immediately), it is over a marginal cost report which implies — through the mechanism [MON] — a quantity.

Given the outcome price and quantity, we can see immediately that

$$\pi_{\text{MON}}(\sigma, C) = \left(\frac{A-C}{2B}\right)\left(\frac{A+C}{2} - \sigma\right)$$

(d) Is $\pi_{\text{MON}}(\sigma, C)$ incentive compatible?

Solution: as before, we need to ensure that $\pi_{\text{MON}}(\sigma,\sigma) \geq \pi_{\text{MON}}(\sigma, C)$ for all C; where in (b) we argued against incentive compatibility by finding a more-optimal report of cost, here we defend incentive compatibility by directly showing that reporting $C = \sigma$ solves the first-order conditions.

From (c), we find the first-order conditions of π_{MON} are

$$\frac{\partial \pi_{\text{MON}}}{\partial C} = \left(\frac{1}{2B}\right)\left(\sigma - \frac{A+C}{2}\right) + \left(\frac{1}{2}\right)\left(\frac{A-C}{2B}\right)$$

Setting the derivative to 0, we find $C = \sigma$. Since truthful reporting maximizes profits, it follows that the [MON] mechanism is incentive compatible.

(e) Apply the Appropriation Principle from efficient, incentive-compatible mechanisms to show that there is a money payment scheme that would induce the seller to report his cost function truthfully.

Solution: here, we are asked to generate a mechanism which is efficient and incentive compatible (the [PTE] mechanism is neither efficient nor incentive compatible; the [MON] mechanism is incentive compatible but still not efficient). The appropriation principle tells us that in order to accomplish this feat, the seller's benefit must be perfectly aligned with social surplus — slightly more formally, the two quantities must be equal except for possibly some constant additive shift.

Since the [MON] mechanism already induces truthful reporting, it would be difficult[2] to add a money payments scheme to derive efficiency; therefore we modify the [PTE] mechanism. Let the money payments scheme be denoted by $m(C)$; from the appropriation principle, we must have

$$\pi_{\text{PTE}}(\sigma, C) + m(C) = g(\sigma, C) = B_s(\sigma, C) + B_b(C)$$

where $g(\cdot,\cdot)$ is the social benefit function, decomposed further into the sum of the seller's surplus and the buyer's surplus *prior to money transfers being issued.*

Notice that the seller's surplus is defined by $B_s(\sigma, C) = \pi_{\text{PTE}}(\sigma, C)$; hence the money transfer scheme must satisfy

$$m(C) = B_b(C) = \int_0^{z(C)} A - Bx - p(C)dx$$

where $z(C)$ is the quantity determined through the [PTE] mechanism, and $p(C)$ is the clearing price. As above, we know that $z(C) = \frac{A-C}{B}$, $p(C) = C$. The seller's benefit is then expressable as

$$\pi_{\text{PTE}}(\sigma, C) + \int_0^{\frac{A-C}{B}} A - Bx - Cdx$$

[2]We may be able to claim, "impossible," but that would require further proof.

To verify incentive compatibility, we take first-order conditions,

$$\begin{aligned}\frac{\partial \pi_{\text{PTE}}}{\partial C} &= \left(\frac{A-C}{B}\right) - \frac{1}{B}(C-\sigma) - \frac{1}{B}\left(A - B\left(\frac{A-C}{B}\right)\right) - \int_0^{\frac{A-C}{B}} dx \\ &= \frac{1}{B}(A - C + \sigma - C - A + A - C + C + C - A) \\ &= \frac{1}{B}(\sigma - C)\end{aligned}$$

Setting this condition to 0, we find that $C = \sigma$ is the optimal report. Further, since the [PTE] mechanism is efficient (under truthful reporting) it follows that this money payment scheme no only induces truthful reporting but also obtains an efficient allocation.

(f) To simplify concerns regarding aggregation, assume there is a single agent in the economy. If this consumer is a strategic agent, can the mechanism derived in (e) be incentive compatible from the consumer's side if budget balance is satisfied?

Solution: if budget balance is satisfied, for the seller to obtain all of the buyer's surplus it must be that the buyer is obtaining 0 surplus. But if A and B are not common knowledge and must be reported by the buyer, the buyer has an incentive to shift his demand curve downward/rightward; by doing so, a lower amount of output is obtained in the economy — an inefficient amount — and a lesser amount of surplus is transferred to the seller. However, since the buyer's *true* demand curve lies higher, he retains some of his surplus.

To demonstrate this algebraically, assume that the seller is truthfully reporting $C = \sigma$. If the buyer submits a demand curve defined by A', B', the [PTE] mechanism generates a market quantity of $z(A', B') = \frac{A'-\sigma}{B'}$. Net of subsequent money transfers to the seller, the buyer's surplus is

$$\int_0^{z(A',B')} A - Bx - C dx - \int_0^{z(A',B')} A' - B'x - C dx = \int_0^{z(A',B')} (A - A') - (B - B')x dx$$

First-order conditions for optimality give us

$$\begin{aligned}\frac{\partial}{\partial A'}: &\qquad 0 = (A - A') - (B - B')z_A(A', B') - \int_0^{z(A',B')} dx \\ \iff &\qquad 0 = (A - A') - (B - B')\frac{1}{B'} - \frac{A' - \sigma}{B'} \\ \iff &\qquad A'\left(1 + \frac{1}{B'}\right) = (B'A - B + B' + \sigma)\frac{1}{B'} \\ \iff &\qquad A' = \frac{B'A - B + B' + \sigma}{B' + 1}\end{aligned}$$

$$\begin{aligned}\frac{\partial}{\partial B'}: &\qquad 0 = (A - A') - (B - B')z_B(A', B') + \int_0^{z(A',B')} x dx \\ \iff &\qquad 0 = (A - A') + \frac{(B - B')(A' - \sigma)}{[B']^2} + \frac{1}{2}\left(\frac{A' - \sigma}{B'}\right)^2 \\ \iff &\qquad 0 = 2(A - A')[B']^2 + 2(B - B')(A' - \sigma) + (A' - \sigma)^2\end{aligned}$$

Of course, this system of equations can be solved; importantly, we can already see that truthful reporting may not be optimal: if $B' = B$, then $A' = \frac{BA+\sigma}{B+1}$. In general, we should not see $A' = A$ at the optimum. Similarly, if $A' = A$, then $B' = \frac{1}{2}(A - \sigma) + B$. As in part (b), truthful reporting is optimal for the agent if and only if $A = \sigma$, which necessarily implies a clearing demand of 0 — in this case surplus in the system, hence there is no surplus to be gained from misreporting.

Interestingly, even if consumers misreport the seller can do no better than to report truthfully, but seeing as we have not solved the consumer's optimum explicitly we will also leave this conclusion to intuition (it is not hard to show graphically; key is that the [PTE] mechanism with transfers awards the seller all apparent surplus).

Resource allocation for a public park

A fixed number of acres of land is to be made into a public park with two possible uses: as sports fields or as botanical gardens. Let $x \in [0, 1]$ be the fraction of the park devoted to sports fields. Assume that no matter what the value of x, costs will be the same, and for purposes of simplification assume these costs are zero. There are N users of the park, each with a utility function

$$v_i(x) + m_i$$

where v_i is strictly concave. Not everyone is a sports enthusiast in the sense that not every v_i is increasing in x,

$$v_i = a_i x - b_i x^2, \quad 0 \leq a_i \leq 2b_i$$

(a) Set up the maximization problem defining an optimal choice of x for $\mathbf{v} = (v_1, \ldots, v_N)$ and use it to derive the first-order conditions for a maximum. Assume that the optimal x is in the interior.

Solution: in the quasilinear model, an optimal/efficient allocation must maximize the sum of non-money utilities, $\sum_{i=1}^N v_i(\cdot)$. We then setup the maximization problem as

$$\max_x \sum_{i=1}^N a_i x - b_i x^2$$

It follows from first-order conditions that the optimal x is such that

$$x = \frac{\sum_{i=1}^N a_i}{2 \sum_{i=1}^N b_i}$$

Assuming that this x is on the interior is equivalent to introducing a single extra qualification on a_i relative to b_i,

$$0 < \sum_{i=1}^N a_i < 2 \sum_{i=1}^N b_i$$

This is a simple refinement of our existing assumption.

(b) Apply the Appropriation Principle for efficient, incentive-compatible mechanisms to demonstrate how money payments can be constructed to induce each individual to report their $v_i \sim (a_i, b_i)$ truthfully.

Solution: the appropriation principle tells us that we can obtain an efficient allocation from an incentive compatible mechanism if each agent's gains follow society's gains, plus or minus some non-strategic constant. Here, this is structured as

$$v_i(x) + m_i = \sum_{j=1}^N a_j x - b_j x^2$$

This immediately gives us a money payment scheme of

$$m_i = \sum_{j \neq i} a_j x - b_j x^2$$

Together, these naturally induce the desired (by taking one algebraic step back) incentive scheme,

$$v_i(x) + m_i = \sum_{j=1}^{N} a_j x - b_j x^2$$

Does this mechanism encourage truthful reporting? Suppose that the x selected by the mechanism is represented by

$$x_i(a,b) = \frac{a + \sum_{j \neq i} a_j}{2b + 2\sum_{j \neq i} b_j}$$

That is, $x_i(a,b)$ is the allocation assigned in the mechanism if agent i reports (a,b) while all other agents $j \neq i$ report truthfully, and the mechanism allocates efficiently based on reports. First-order conditions with respect to agent i's report are then

$$\begin{aligned}
\frac{\partial}{\partial a}: \qquad & 0 = \sum_{j=1}^{N} a_j [x_i]_a(a,b) - 2b_j [x_i]_a(a,b) x_i(a,b) \\
\iff \qquad & 0 = \sum_{j=1}^{N} a_j - 2b_j x_i(a,b) \\
\iff \qquad & x_i(a,b) = \frac{\sum_{j=1}^{N} a_j}{2\sum_{j=1}^{N} b_j}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial}{\partial b}: \qquad & 0 = \sum_{j=1}^{N} a_j [x_i]_b(a,b) - 2b_j [x_i]_b(a,b) x_i(a,b) \\
\iff \qquad & 0 = \sum_{j=1}^{N} a_j - 2b_j x_i(a,b) \\
\iff \qquad & x_i(a,b) = \frac{\sum_{j=1}^{N} a_j}{2\sum_{j=1}^{N} b_j}
\end{aligned}$$

Following from the definition of $x_i(a,b)$, it follows that truthful reporting $(a,b) = (a_i, b_i)$ is *an* optimal outcome; more generally, any report of (a,b) which does not alter the ratio of the sums of the coefficients will technically be optimal. However, since truthful reporting is at least as good as any other report, we see that we have an incentive-compatible mechanism which by construction generates the efficient outcome.

(c) Show that if the money payments schedule to individual i in (b) is modified to include a term that is allowed to vary with the utility of others (i.e., v_{-i}) but is constant with respect to i's reported utility, the conclusions in (b) remain the same. Call this constant $c_i(v_{-i})$.

Solution: let $m_i(\mathbf{v})$ denote the money transfer scheme m_i from part (b) above. We will now include an additional term in the transfer,

$$m_i = m_i(\mathbf{v}) + c_i(v_{-i})$$

The agent's utility is now given by

$$v_i(x) + m_i(\mathbf{v}) + c_i(v_{-i}) = c_i(v_{-i}) + \sum_{j=1}^{N} a_j x - b_j x^2$$

Taking first-order conditions, since the $c_i(\cdot)$ term depends only on the reports of others — not on their final utility, merely their reports — and not on the report of i, the $c_i(\cdot)$ term falls out and will not distort the conditions for optimal reporting by agent i. It follows that the introduction of this term does not affect the incentive compatibility of the mechanism from (b).

(d) Let $c_i(v_{-i})$ be the maximum total gains to the model $\mathbf{v}_{-i} = (v_j)_{j \neq i}$, the economy without agent i. Find the expression for $m_i(\mathbf{v})$ when the money payments function satisfies the conditions in (b).

Solution: appealing to part (a), first-order conditions in this model reveal an efficient allocation of

$$x_{-i} = \frac{\sum_{j \neq i} a_j}{2 \sum_{j \neq i} b_j}$$

Total gains in this system are

$$\sum_{j \neq i} a_j \left(\frac{\sum_{k \neq i} a_k}{2 \sum_{k \neq i} b_k} \right) - b_j \left(\frac{\sum_{k \neq i} a_k}{2 \sum_{k \neq i} b_k} \right)^2 = \frac{\sum_{j \neq i} \sum_{k \neq i} a_j a_k}{4 \sum_{j \neq i} b_j}$$

By our desired construction, this is the term $-c_i(\mathbf{v}_{-i})$ (there was no claimed sign convention to $c_i(\cdot)$, so a positive transfer was assumed in (c); it is evident that here we prefer a payment back to the designer).

The total transfers to agent i are then

$$\begin{aligned} m_i(\mathbf{v}) &= \sum_{j \neq i} a_j x^* - b_j [x^*]^2 - \frac{\sum_{j \neq i} \sum_{k \neq i} a_j a_k}{4 \sum_{j \neq i} b_j} \\ &= \sum_{j=1}^{N} a_j x^* - b_j [x^*]^2 - \frac{\sum_{j \neq i} \sum_{k \neq i} a_j a_k}{4 \sum_{j \neq i} b_j} - a_i x^* + b_j [x^*]^2 \\ &= \frac{\sum_{j=1}^{N} \sum_{k=1}^{N} a_j a_k}{4 \sum_{j=1}^{N} b_j} - \frac{\sum_{j \neq i} \sum_{k \neq i} a_j a_k}{4 \sum_{j \neq i} b_j} - a_i x^* + b_i [x^*]^2 \end{aligned}$$

(e) With the constant term $c_i(v_{-i})$ in (d), do the money payments sum to zero, less than zero, or greater than zero? Can you explain why in terms of the language of externalitites?

Solution: the introduction of a new agent will, in general, alter the efficient allocation. But if the efficient allocation is altered, the society without the new agent must have been better off before. It follows that the general case is that the introduction of a new agent must reduce the welfare of the society without the agent. Intuitively, then, the sum of the transfers should be negative, with an exception for the case in which all agents share a set of preferences where the sum of transfers should be zero.

Algebraically, this is an absolute nightmare to demonstrate; however, we can use the language of conditional optimization above to construct a reasonably mathematical argument with no explicit calculations. Suppose that x^* is the optimal allocation when i is in the economy, and x^*_{-i} is the optimal allocation when i is not in the economy. This tells us that

$$\sum_{j \neq i} a_j x^*_{-i} - b_j [x^*_{-i}]^2 \geq \sum_{j \neq i} a_j x - b_j x^2$$

for all x; in particular,

$$\sum_{j \neq i} a_j x^*_{-i} - b_j [x^*_{-i}]^2 \geq \sum_{j \neq i} a_j x^* - b_j [x^*]^2$$

It follows by the definition of the money transfer function that

$$m_i(\mathbf{v}) = \left(\sum_{j \neq i} a_j x^* - b_j [x^*]^2 \right) - \left(\sum_{j \neq i} a_j x^*_{-i} - b_j [x^*_{-i}]^2 \right) \leq 0$$

Since the money payment to any individual agent is weakly negative, it follows that the sum of payments should be weakly negative, as well.

Note that this result is in direct opposition to many of our familiar public goods problems (particularly those from Riley's section of the course). The difference here is that the introduction of an additional agent exerts a negative externality on all other agents; in Riley's traditional setup, the introduction of an additional agent makes production of the public good more likely to be efficient, hence raising not only social welfare but also each individual's welfare[3].

(f) Suppose that agents have the option to not fund the sports fields, but in doing so they will be exiled for eternity. Agents value exile idiosyncratically at $-r_i$; how does this change the above conclusions?

Solution: we check that the mechanism proposed above satisfies the necessary participation constraints with outside options given by $-r_i$. By construction of the money transfer function, agent utility is

$$\begin{aligned}
u_i &= \frac{\sum_{j=1}^N \sum_{k=1}^N a_j a_k}{4\sum_{j=1}^N b_j} - \frac{\sum_{j\neq i}\sum_{k\neq i} a_j a_k}{4\sum_{j\neq i} b_j} \\
&= \left[4\left(\sum_{j=1}^N b_j\right)\left(\sum_{j\neq i} b_j\right)\right]^{-1}\left[\sum_{j=1}^N\sum_{k=1}^N\sum_{h\neq i} a_j a_k b_h - \sum_{j\neq i}\sum_{k\neq i}\sum_{h=1}^N a_j a_k b_h\right] \\
&= \left[4\left(\sum_{j=1}^N b_j\right)\left(\sum_{j\neq i} b_j\right)\right]^{-1}\left[\left(2\sum_{j=1}^N a_i a_j - a_i^2\right)\sum_{h=1}^N b_h - b_i\sum_{j=1}^N\sum_{k=1}^N a_j a_k\right] \\
&= \left[4\left(\sum_{j=1}^N b_j\right)\left(\sum_{j\neq i} b_j\right)\right]^{-1}\left[2\sum_{j\neq i}\sum_{h\neq i} a_i a_j b_h - \sum_{j\neq i}\sum_{k\neq i} a_j a_k b_i\right] \\
&= \left[4\left(\sum_{j=1}^N b_j\right)\left(\sum_{j\neq i} b_j\right)\right]^{-1}\left[\sum_{j\neq i} a_j\left(\sum_{h\neq i} 2a_i b_h - a_h b_i\right)\right]
\end{aligned}$$

If this quantity is less than $-r_i$, there will problems with the participation constraint[4]. Since further derivation appears to quickly grow frustrating, we'll focus on brief intuition.

It $u_i \geq -r_i$, everything is copacetic and the mechanism is still efficient and incentive compatible. Otherwise, an efficient mechanism must allow the agent to exercise his outside option; if exercising the outside option is advantageous, optimal social surplus is then the efficient social surplus when agent i is not within the economy plus the value of the agent exercising his outside option. When the social surplus of the economy without i is subtracted, we find that the agent's payoff from the mechanism is equivalent to his payoff from the outside option; that is, the efficient mechanism with transfers forces the agent to leave the mechanism! This is consistent with efficiency, and the outcome perfectly matches our expectations.

2006 Spring comp, question 5

Individuals face two kinds of choices. In one, the choices are PUB $= \{0, 1\}$ in which the color of the asphalt on the streets will remain what it has been (0) or be painted to light green (1). In the second, the choices are PRI $= \{s_1, \ldots, s_N\}$, where $s_i \in \{0, 1\}$ and $\sum_i s_i = 1$; $s_i = 1$ means that i gets a two-week (expenses

[3]For a very similar argument that takes a slightly different tack, see the solution to part (d) of the following question.

[4]Excuse algebra errors in this derivation — and email me with corrections — as sums quickly become unwieldy.

paid) trip to Tokyo and everyone else stays home. Normalize utility to be 0 whenever s or s_i is 0. Individual i's tastes are determined by the number w_i according to:

- $v_i(s) = w_i$, if $s = 1$ and $v_i(s) = 0$ otherwise, when $s \in$ PUB.
- $v_i(s) = w_i$, if $s = s_i = 1$, and $v_i(s) = 0$ otherwise, when $s \in$ PRI.

Hence, the characteristics of the economy are given by a vector $w = (w_1, \ldots, w_N)$ whether the choices are in PUB or PRI. For choices in PUB, allow the possibility that $w_i < 0$, whereas for PRI assume $w_i \geq 0$. There are no cost differentials between choices in PUB or among those in PRI, so treat costs as zero.

(a) What are the conditions for an optimal choice for w in PUB? In PRI?

Solution: with respect to the public good, the efficient allocation is $s = 1$ when $\sum_{i=1}^{N} w_i \geq 0$ and $s = 0$ otherwise (ignoring the nuances of what defines efficiency when the sum is exactly 0). With respect to the private good allocation problem, the notion of optimality depends crucially on the ability to enact money transfers between agents; without such transferrable utility, any assignment of the private good to agent i with $w_i > 0$ is efficient. With transferrable utility (which we will assume henceforth) the optimal assignment is to the agent i with $w_i > w_j$ for all $j \neq i$ (with the required random assignment in the event of two agents with identically-high valuations).

A *mechanism* can be regarded as a mapping from economies w to $(s(w), p_1(w), \ldots, p_N(w))$, where $s(w) \in$ PUB or PRI and $p_i(w)$ is the money payment made by i. The utility of the outcome to i is $v_i(s(w)) - p_i(w)$ assuming everyone reports honestly, i.e. $v_i(s = 1) = w_i$.

(b) Define the conditions for a mechanism to encourage individuals to truthfully reveal their preferences. Are the conditions for PUB different from PRI? (ignore incentive compatibility conditions with respect to the supply of PUB or PRI)

Solution: truthful revelation is equivalent to incentive compatibility. The incentive compatibility criterion requires

$$v_i(s(w)) - p_i(w) \geq v_i(s(w', w_{-i})) - p_i(w', w_{-i})$$

for all feasible reports w'. Notice that the requirement that truthful reporting — and the mathematical statement expressing this requirement — does not change between PUB and PRI; that is to say, in either case we only require that each agent is best-responding with respect to *only her own* incentives and preferences, and in no case is she necessarily acting intentionally for social welfare.

(c) Suppose a mechanism satisfies the optimality conditions of (a). Give sufficient conditions for the mechanism to satisfy the incentive compatibility condition of (b). Outline an argument to justify your claim. Do the conditions vary between PUB and PRI?

Solution: following the appropriation principle, a sufficient condition for incentive compatibility in an efficient mechanism is that an agent is rewarded with her effect on social welfare rather than simply her effect on her own welfare. Intuitively, when this holds the agent is properly incentivized to maximize social welfare (to the extent she can through her own report); when each agent is faced with this incentive scheme, the efficient outcome is necessarily obtained. This condition does not change when it is applied to a public-goods situation versus a private-goods situation; in either case, rewarding based on social outcomes rather than individual outcomes will yield efficiency.

Mathematically, this form is often given by

$$v_i(s(w', w_{-i})) - p_i(w', w_{-i}) = \sum_{j=1}^{N} v_j(s(w', w_{-i})) - \sum_{j \neq i} v_j(s_{-i}(w_{-i}))$$

Assuming an efficient allocation rule, this yields a payment function of

$$p_i(w) = \sum_{j \neq i} v_j(s_{-i}(w_{-i})) - v_j(s(w))$$

Of course, this can be modified by some additive payment which is independent of the agent's reported type, $c_i(w_{-i})$, if necessary. In the formulation presented in the lecture notes, $c_i(w_{-i})$ is a constant independent of all parameters; the more general form stated here is equally valid and slightly more general. If it helps, feel free to assume $c_i(w_{-i}) = k$ for some constant k; most applications we see will require nothing more complex than this setup.

Two conditions on a mechanism are: *weak budget balance*, for all w, $\sum_i p_i(w) \geq 0$; and, *voluntarism*, for all w, $v_i(s(w)) - p_i(w) \geq 0$.

(d) Do the conditions for an optimal incentive compatible mechanism in (c) imply differences with respect to weak budget balancing and voluntarism for PUB compared to PRI? Explain.

Solution: we will assume the voluntarism (participation) constraint is satisfied.

In the case of PRI, voluntarism tells us that $c_i(w_{-i}) \leq 0$ (where $c_i(w_{-i})$ is as in part (c) above); otherwise, an agent with the lowest possible type, $w_i = 0$, expects to make some payment to the designer while expecting 0 payoff, a violation of the voluntarism constraint. For budget balance, it is natural to let $c_i(w_{-i}) \geq 0$ to maximize designer payoff; putting these two implications together we obtain $c_i(w_{-i}) = 0$. Further, for all agents other than those with the highest w_i, $\sum_{j \neq i} v_j(s_{-i}(w_{-i})) = \sum_{j \neq i} v_j(s(w))$; that is, removing the agent from the economy does not affect the efficient allocation and hence does not affect societal gains through the mechanism. For the agent with the highest w_i, we know $\sum_{j \neq i} v_j(s(w)) = 0$ — no other agent receives any positive surplus in the efficient allocation of the good to agent i — but $\sum_{j \neq i} v_j(s_{-i}(w_{-i})) \geq 0$, since if agent i is not present an efficient allocation must give the good to some other agent j, who by assumption has value $w_j \geq 0$. It follows that $p_i(w) \geq 0$ for this agent. Hence $\sum_{i=1}^{N} p_i(w) \geq 0$ for all w, and in PRI we can simultaneously satisfy voluntarism and weak budget balance.

In the case of PUB, voluntarism still tells us that $c_i(w_{-i}) \leq 0$; consider, for example, the case in which w_i is uniformly 0 for each agent. Then $v_i(s(w)) = 0$, and hence $p_i(w) \leq 0$ for all i. Given the construction of $p_i(\cdot)$, it is immediate that $c_i(w_{-i}) \leq 0$ in this particular case. Since $c_i(w_{-i})$ is independent of the agent's type, we must generally have $c_i(w_{-i}) \leq 0$ for voluntarism[5].

However, weak budget balance may possibly be more of a problem. Intuitively, it is difficult to incentivize agents not to free ride on each others' reports, and the need to incentivize each agent to do so will come into conflict with the voluntarism constraint that $c_i(w_{-i}) \leq 0$. The sum of payments is

$$\sum_{i=1}^{N} p_i(w) = \sum_{i=1}^{N} \left[\sum_{j \neq i} v_j(s_{-i}(w_{-i})) - v_j(s(w)) + c_i(w_{-i}) \right]$$

Three cases arise:

- The introduction of agent i has no effect on the socially efficient allocation. Then we have

$$\sum_{j \neq i} v_j(s_{-i}(w_{-i})) = \sum_{j \neq i} v_j(s(w))$$

and the payment is equal to $c_i(w_{-i})$.

[5] When we talk of issuing constant transfers to help balance the budget, we do not allow agents to leave the mechanism; the IRS in this case essentially will not allow you to move to Canada. Of course, this can probably be supported with voluntarism against some particularly negative outside option — say a deathly fear of lumberjacks — but there is no need to overcomplicate matters in this immediate setup.

- The introduction of agent i manipulates the social preferences so that $s = 1$ is no longer the efficient allocation. Then $\sum_{j\neq i} v_j(s_{-i}(w_{-i})) > 0$, yet $\sum_{j\neq i} v_j(s(w)) = 0$; the payment is therefore positive for some $c_i(w_{-i}) \leq 0$.

- The introduction of agent i manipulates the social preferences so that $s = 0$ is no longer the efficient allocation. Then $\sum_{j\neq i} v_j(s_{-i}(w_{-i})) = 0$, but $\sum_{j\neq i} v_j(s(w)) > -w_i$; further, $\sum_{j\neq i} v_j(1) < 0$, as otherwise $s = 1$ would be the efficient allocation without agent i in the economy. It follows again that the payment is positive for some $c_i(w_{-i}) \leq 0$.

In each of the above cases, the money transfer is weakly positive for $c_i(w_{-i}) = 0$; this roughly matches intuition, since if agent i alters the social allocation in PUB her type report is, in effect, moving the rest of society to an allocation which is suboptimal absent her input. It follows that if she is to be rewarded on the basis of her contribution to social surplus, she should pay into the mechanism.

Immediately, we see that weak budget balance can obtain with voluntarism in PUB. How does this differ from the context in which we are attempting to model trade between a buyer (positive w_b) and seller (negative w_s), where we have seen that a weakly positive injection of money is nnecessary? We need to be careful about making such a comparison: in this model, the existence of a single buyer in the system with no seller would actually cause the public good to be allocated, where in the buyer-seller setup such an outcome is nonsensical. So where in the standard buyer-seller setup, the seller necessarily contributes a weakly positive amount to social surplus, in this setting the "seller" can effectively cut the buyer's surplus by indicating a significantly low type. Thus the analogy to the previously-analyzed buyer-seller case is imperfect, and hence there is no inconsistency between the positive contributions to the designer in this case and the positive contributions to the agents in the known buyer-seller case.