

Incentivos y la asignación de bienes públicos

Nota de clase. Guía de lectura de J.-J. Laffont (1987),
Handbook of Public Economics, vol. II, cap. 10

Profesor: José Gallardo Ku (j.gallardo@pucp.edu.pe)
Jefes de práctica: Marcelo Gallardo (marcelo.gallardo@pucp.edu.pe),
Raúl Amao (raul.amao@pucp.edu.pe)

1 Introducción

Consideramos una economía con I individuos, llamados *agentes*, indexados por $i = 1, 2, \dots, I$. Aquí I es la cantidad de personas que participan de la decisión.

Un *bien público* es un bien que, una vez provisto, todos consumen sin rivalidad ni exclusión: el alumbrado de una calle, la defensa nacional, un parque. La cantidad eficiente y de ese bien satisface la condición de Samuelson,

$$\sum_{i=1}^I \text{TMS}_{y,x}^i = \text{TMT}_{y,x},$$

es decir, la suma sobre los I agentes de sus tasas marginales de sustitución entre el bien público y y el privado x (cuánto bien privado cede cada uno por una unidad más de y) iguala la tasa marginal de transformación (el costo de producir una unidad más de y).

La dificultad es de información. Para aplicar la condición de Samuelson, quien decide necesita las tasas marginales *verdaderas* de cada agente, pero esa información es *privada*: cada agente conoce la suya y nadie más la observa. Además, cada agente tiene incentivo a declararla por debajo de su valor verdadero, para que el bien se financie con aportes ajenos. Este es el problema del polizón (*free-rider*). La pregunta del capítulo es si puede diseñarse una regla de decisión que induzca a los agentes a revelar su información verdadera y que, al hacerlo, produzca un resultado eficiente.

2 ¿Qué es un mecanismo?

2.1 Un ejemplo concreto

Dos vecinos, Johel y Jorge, comparten una calle y deben decidir si instalan un farol. El farol es un bien público: si se instala, ambos lo disfrutan; si no, ninguno. Instalarlo cuesta 100 soles. Johel valora el farol en una cantidad θ^A (en soles) y Jorge en θ^B ; cada uno conoce su propia valoración pero no la del otro. La decisión eficiente es instalar el farol si y solo si el beneficio total no es menor que el costo,

$$\text{instalar} \iff \theta^A + \theta^B \geq 100.$$

El planificador (la junta de vecinos, la municipalidad) no observa θ^A ni θ^B . Si pregunta directamente “¿cuánto vale el farol para ti?”, cada vecino responde estratégicamente: según cómo se reparta el costo, le conviene exagerar (para que se instale a costa del otro) o minimizar (para pagar menos). Una pregunta directa no basta; hace falta una regla bien diseñada.

2.2 Los ingredientes

Definición 1 (Mensaje). Un *mensaje* de un agente es lo que ese agente comunica al planificador. En el ejemplo, el mensaje de Johel podría ser una cifra m^A (“declaro que el farol vale m^A para mí”). El conjunto de mensajes que puede enviar es su *espacio de mensajes* M^i .

El punto central es que el agente *elige* su mensaje; no está obligado a decir la verdad, y puede declarar $m^A \neq \theta^A$ si le conviene.

Definición 2 (Función de resultado). Una *función de resultado* es una regla, fijada y anunciada de antemano, que asocia a cada combinación de mensajes una decisión completa. Se denota g . En el ejemplo, g recibe el par (m^A, m^B) y devuelve si el farol se instala y cuánto paga cada vecino.

Por ejemplo, una función de resultado posible es: “instalar si $m^A + m^B \geq 100$ y cobrar 50 a cada uno; si no, no instalar ni cobrar”. Cada g define un problema de decisión distinto para los agentes.

Definición 3 (Mecanismo). Un *mecanismo* es el par (M, g) : los espacios de mensajes $M = M^1 \times \dots \times M^I$ y la función de resultado g . Es el conjunto de reglas que el planificador se compromete a seguir; el agente conoce g antes de elegir su mensaje.

Una vez fijado el mecanismo, cada agente enfrenta una pregunta: “¿qué mensaje envío?”. Como el resultado depende también de lo que envíen los demás y cada uno busca el mejor resultado para sí, esta situación es un *juego*: una situación en la que varias personas eligen acciones y el resultado de cada una depende de lo que eligen todas.

Definición 4 (Estrategia). Una *estrategia* de un agente es la regla con que elige su mensaje; puede depender de su información privada. Formalmente, es una función $s^i : \Theta^i \rightarrow M^i$ que a cada tipo $\theta^i \in \Theta^i$ le asigna el mensaje $s^i(\theta^i) \in M^i$ que enviará. La estrategia de Johel es la función que a cada valoración posible θ^A le asigna su mensaje m^A .

2.3 ¿Cómo predecir qué harán los agentes?

Para evaluar un mecanismo necesitamos predecir el comportamiento: eso es un *concepto de solución*, una hipótesis sobre cómo juegan los agentes. El más exigente y deseable es el siguiente.

Definición 5 (Estrategia dominante). Una estrategia es *dominante* para un agente si es la mejor que puede elegir sin importar lo que hagan los demás: cualquiera sea el mensaje del otro vecino, a Ana le conviene al menos tanto su estrategia dominante como cualquier otra.

La virtud de una estrategia dominante es que el agente no necesita adivinar nada sobre los demás ni coordinarse: le conviene jugarla en todos los casos. Por eso, cuando un mecanismo hace que decir la verdad sea dominante, podemos confiar en que los agentes dirán la verdad. Sin embargo, más adelante veremos que la estrategia dominante puede ser demasiado exigente, y recurriremos a conceptos más débiles.

2.4 El objetivo: implementar una regla

El planificador tiene una meta: una *función de elección social* f , que a cada perfil de valoraciones verdaderas le asigna el resultado que querría *si* conociera esa información. En el ejemplo, f es

“instalar si $\theta^A + \theta^B \geq 100$ ”. No puede aplicar f directamente porque no observa las valoraciones; lo que puede hacer es diseñar un mecanismo y esperar que el comportamiento de los agentes reproduzca f .

Definición 6 (Implementación). Un mecanismo (M, g) *implementa* la función de elección social f si, cuando los agentes juegan según el concepto de solución, el resultado coincide con f para todo perfil de valoraciones verdaderas. El mecanismo es *directo* si los mensajes son las propias valoraciones ($M^i = \Theta^i$) y *revelador* (veraz) si, en ese mecanismo directo, decir la verdad es la estrategia de equilibrio.

Con esto cerramos el vocabulario. El resto del capítulo estudia qué funciones f admiten un mecanismo que las implemente y cómo construirlo.

2.5 Notación y formalización

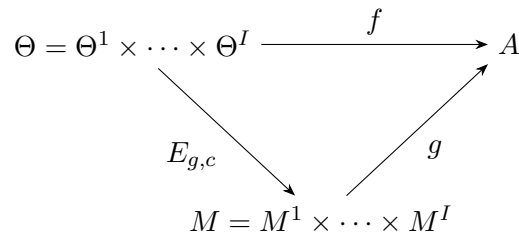
Para un agente i , su valoración verdadera (su *tipo*) es $\theta^i \in \Theta^i$; el perfil de todos los tipos es $\theta = (\theta^1, \dots, \theta^I)$ y el de los demás es $\theta^{-i} = (\theta^1, \dots, \theta^{i-1}, \theta^{i+1}, \dots, \theta^I)$.

- I número de agentes; $i \in \{1, \dots, I\}$.
- $\theta^i \in \Theta^i$ tipo (valoración) de i , información privada; θ^{-i} : tipos de los demás.
- $f : \Theta \rightarrow A$ función de elección social; A es el conjunto de resultados.
- (M, g) mecanismo: mensajes $M = M^1 \times \dots \times M^I$, función de resultado $g : M \rightarrow A$.
- \succeq_{θ^i} preferencia de i cuando su tipo es θ^i ; $u^i(a, \theta^i)$ su representación cardinal.
- $\psi^i(\theta^{-i})$ creencias de i sobre los tipos de los demás (cuando se usen).
- $v^i(y, \theta^i), t^i$ valoración del bien público y transferencia monetaria (caso cuasi-lineal $u^i = x^i + v^i$).

Fijado un concepto de solución, sea $E_{g,c}(\theta)$ el conjunto de mensajes de equilibrio cuando el perfil verdadero es θ . El mecanismo *implementa* a f si

$$g(E_{g,c}(\theta)) = f(\theta), \quad \forall \theta \in \Theta.$$

Esto se resume en el diagrama de Mount–Reiter¹: el camino directo f debe coincidir con el indirecto (jugar el equilibrio $E_{g,c}$ y aplicar g).



¹El triángulo conmutativo de Mount y Reiter (1974) expresa que los dos caminos de Θ a A —el directo f y el indirecto $g \circ E_{g,c}$ — coinciden.

3 Conceptos de solución

De la noción más fuerte a la más débil (sea m^{-i} el perfil de mensajes de los demás):

- *Estrategia dominante*: $g(m^{*i}, m^{-i}) \succeq_{\theta^i} g(m^i, m^{-i})$ para todo m^i y todo m^{-i} .

- *Bayes–Nash*: con $m^*(\theta) = (m^{*1}(\theta^1), \dots, m^{*I}(\theta^I))$,

$$\int_{\Theta^{-i}} u^i(g(m^*(\theta)), \theta^i) \psi^i(\theta^{-i}) d\theta^{-i} \geq \int_{\Theta^{-i}} u^i(g(m^i, m^{*-i}), \theta^i) \psi^i(\theta^{-i}) d\theta^{-i}, \quad \forall m^i.$$

- *Maximin*: $\min_{m^{-i}} u^i(g(m^{*i}, m^{-i}), \theta^i) \geq \min_{m^{-i}} u^i(g(m^i, m^{-i}), \theta^i)$.

- *Nash*: $g(m^{*i}, m^{*-i}) \succeq_{\theta^i} g(m^i, m^{*-i})$.

4 El principio de revelación

Teorema 1 (Gibbard, 1973). *Si algún mecanismo implementa f en estrategias dominantes, entonces el mecanismo directo y veraz también implementa f en estrategias dominantes. En consecuencia, basta estudiar mecanismos directos en los que decir la verdad es dominante.*

Demostración. Sea (M, g) un mecanismo que implementa f en estrategias dominantes.

Paso 1. Como cada agente tiene una estrategia dominante, para cada tipo θ^i existe un mensaje $m^{*i}(\theta^i)$ óptimo para i cualquiera sea el mensaje de los demás. Que el mecanismo implemente f significa que, al jugar todos su mensaje dominante,

$$g(m^{*1}(\theta^1), \dots, m^{*I}(\theta^I)) = f(\theta), \quad \forall \theta.$$

Paso 2. Definimos el mecanismo directo (Θ, ψ) : cada agente declara un tipo $\hat{\theta}^i$, y

$$\psi(\hat{\theta}) \equiv g(m^{*1}(\hat{\theta}^1), \dots, m^{*I}(\hat{\theta}^I)).$$

Por el Paso 1, $\psi(\theta) = f(\theta)$ para todo θ : el mecanismo directo produce f cuando todos declaran la verdad.

Paso 3. Veamos que a cada agente le conviene declarar su tipo verdadero. Fijemos i con tipo verdadero θ^i y supongamos que los demás declaran tipos cualesquiera $\hat{\theta}^{-i}$. Si i declara la verdad, el resultado es $g(m^{*i}(\theta^i), m^{*-i}(\hat{\theta}^{-i}))$; si miente declarando $\tilde{\theta}^i$, es $g(m^{*i}(\tilde{\theta}^i), m^{*-i}(\hat{\theta}^{-i}))$. En el mecanismo original, $m^{*i}(\theta^i)$ es dominante para el tipo θ^i , de modo que frente al perfil $m^{*-i}(\hat{\theta}^{-i})$ da un resultado al menos tan bueno como cualquier otro mensaje, en particular $m^{*i}(\tilde{\theta}^i)$. Por tanto

$$g(m^{*i}(\theta^i), m^{*-i}(\hat{\theta}^{-i})) \succeq_{\theta^i} g(m^{*i}(\tilde{\theta}^i), m^{*-i}(\hat{\theta}^{-i})),$$

es decir, declarar la verdad es óptimo cualesquiera sean las declaraciones de los demás. ■

5 La imposibilidad de Gibbard–Satterthwaite

Teorema 2 (Gibbard, 1973; Satterthwaite, 1975). *Si el rango de f tiene al menos tres resultados, el dominio es universal (todo perfil de preferencias estrictas es admisible) y f es implementable en estrategias dominantes, entonces f es dictatorial: existe un agente cuyo resultado preferido es siempre el elegido.*

Intuición. Hay un conjunto finito de **alternativas** (los posibles resultados) $A = \{a_1, \dots, a_m\}$ con $m \geq 3$, y n agentes. Cada agente i reporta un **orden total estricto** \succ_i de *todas* las alternativas de A —su ranking, de la mejor a la peor—, y “ x está **primero** para i ” significa que x es la cima de ese ranking: $x \succ_i y$ para toda $y \neq x$. La regla f toma el perfil $(\succ_1, \dots, \succ_n)$ y devuelve *una* alternativa de A . El argumento solo necesita tres alternativas; para verlo con claridad fijemos $A = \{a, b, c\}$.

El votante pivote. Una regla con monotonía y unanimidad no puede repartir la decisión: existe un único agente decisivo. Para encontrarlo, construimos una cadena de $n+1$ perfiles en la que la alternativa b sube del último al primer lugar, *un agente a la vez*, manteniendo fijo el orden relativo de las otras dos ($a \succ c$). Usamos la notación compacta “ acb ” = el ranking $a \succ c \succ b$. Así, cada agente tiene a b en el fondo (acb) o en la cima (bac):

perfil	ag.1	ag.2	...	ag. n	f
P_0	acb	acb	...	acb	a
P_1	bac	acb	...	acb	\cdot
\vdots					\vdots
P_{i^*-1}	bac	... bac	acb ...	acb	$\neq b$
P_{i^*}	bac	... bac bac	acb ...	acb	$= b$
\vdots					\vdots
P_n	bac	bac	...	bac	b

En P_k , los agentes $1, \dots, k$ tienen a b en la cima y el resto la tiene en el fondo. Los dos extremos están forzados:

- en P_0 *todos* ponen a a primero, así que por unanimidad $f(P_0) = a$, en particular $f(P_0) \neq b$;
- en P_n *todos* ponen a b primero, así que por unanimidad $f(P_n) = b$.

El resultado arranca en a y termina en b : en algún paso cambia a b por primera vez. Sea i^* ese paso. Lo único que ocurre entre P_{i^*-1} y P_{i^*} es que el agente i^* levanta a b del fondo a la cima —nadie más se mueve—, de modo que el salto del resultado a b es atribuible *exclusivamente* a él. Ese agente es el **pivote**.

El pivote no solo decide sobre b . Aquí entra la tercera alternativa c : variando con cuál alternativa se “pivota”, dos lemas estándar de propagación —que usan monotonía y la existencia de c — extienden la conclusión a que ese mismo agente i^* decide el resultado entre *cualquier* par de alternativas y en *cualquier* perfil: lo que i^* pone en primer lugar es siempre lo elegido. Ese agente es el dictador.

La moraleja es constructiva: para obtener reglas no dictatoriales hay exactamente dos salidas,

- restringir el dominio de preferencias admisibles, o
- admitir transferencias monetarias y/o debilitar el concepto de solución.

6 Salida A: votante mediano

Proposición 1 (Dummett–Farquharson, 1961; Moulin, 1980). *Si los resultados se ordenan en una recta y cada agente tiene un único óptimo (su pico p_i) con utilidad que decrece al alejarse, entonces, con I impar, elegir el pico del agente mediano es a prueba de estrategias, eficiente y anónimo. No requiere transferencias.*

Demostración. El agente mediano no gana mintiendo: cualquier pico distinto del suyo solo aleja el resultado de su óptimo o lo deja igual. Un agente con su pico a la derecha de la mediana no cambia nada declarando otro pico también a la derecha, y si declara uno a la izquierda mueve el resultado en sentido contrario al que desea; simétricamente para los de la izquierda. Luego declarar el pico verdadero es dominante. ■

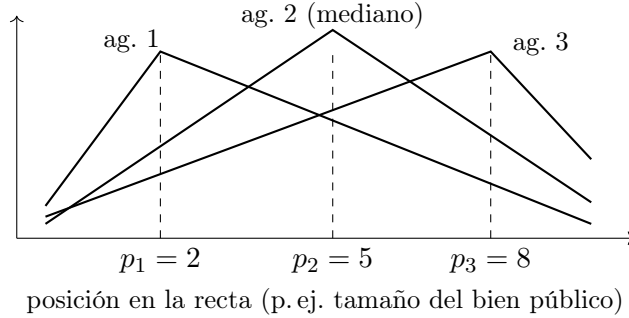


Figura 1: Tres preferencias unimodales; la votación por mayoría selecciona el pico mediano $y^* = 5$.

Ejemplo 1. Picos $p_1 = 2$, $p_2 = 5$, $p_3 = 8$; la regla elige la mediana, $y^* = 5$. Si el agente 3 declara 10, la mediana de $\{2, 5, 10\}$ sigue siendo 5; si declara 4, la mediana de $\{2, 4, 5\}$ es 4, más lejos de su pico. Declarar la verdad es óptimo.

7 El enfoque diferencial

Es el método con el que se construyen mecanismos veraces sin revisar, una por una, todas las declaraciones falsas: se usa la condición de primer orden del agente, se la convierte en identidad y se integra. Lo presentamos en una economía de intercambio con dos bienes y utilidad lineal en el parámetro desconocido,

$$U^i(x_1^i, x_2^i; \hat{\theta}^i) = \hat{\theta}^i x_2^i - \frac{(x_2^i)^2}{2} + x_1^i,$$

donde $\hat{\theta}^i$ es la característica verdadera (privada) y θ^i la que declara. El mecanismo asigna $x_1^i(\theta^i, \theta^{-i})$ y $x_2^i(\theta^i, \theta^{-i})$. Restringimos a mecanismos diferenciables.

7.1 Condición de primer orden y la identidad

Paso 1. (El problema del agente.) El agente elige qué declarar para maximizar su utilidad verdadera:

$$\max_{\theta^i} \left\{ \hat{\theta}^i x_2^i(\theta^i, \theta^{-i}) - \frac{1}{2} x_2^i(\theta^i, \theta^{-i})^2 + x_1^i(\theta^i, \theta^{-i}) \right\}.$$

Paso 2. (Condición de primer orden.) Derivando respecto de la variable de elección θ^i e igualando a cero,

$$\hat{\theta}^i \frac{\partial x_2^i}{\partial \theta^i} - x_2^i \frac{\partial x_2^i}{\partial \theta^i} + \frac{\partial x_1^i}{\partial \theta^i} = 0.$$

Paso 3. (Volverla identidad.) Para que la solución sea decir la verdad, $\theta^i = \hat{\theta}^i$, y que valga para todo $\hat{\theta}^i$, se pide que sobre la diagonal la condición de primer orden sea una identidad en θ^i :

$$\theta^i \frac{\partial x_2^i}{\partial \theta^i}(\theta^i, \theta^{-i}) - x_2^i(\theta^i, \theta^{-i}) \frac{\partial x_2^i}{\partial \theta^i}(\theta^i, \theta^{-i}) + \frac{\partial x_1^i}{\partial \theta^i}(\theta^i, \theta^{-i}) = 0. \quad (5.2)$$

7.2 Integración: la forma del mecanismo

Paso 1. (Despejar.) De (5.2), agrupando los dos primeros términos,

$$\frac{\partial x_1^i}{\partial \theta^i} = (x_2^i - \theta^i) \frac{\partial x_2^i}{\partial \theta^i}.$$

Paso 2. (Integrar de 0 a θ^i .)

$$x_1^i(\theta^i, \theta^{-i}) - x_1^i(0, \theta^{-i}) = \int_0^{\theta^i} (x_2^i(s^i, \theta^{-i}) - s^i) \frac{\partial x_2^i}{\partial s^i}(s^i, \theta^{-i}) ds^i.$$

Paso 3. (Nombrar la constante de integración.) Como $x_1^i(0, \theta^{-i})$ no depende de θ^i , lo llamamos $h^i(\theta^{-i})$:

$$x_1^i(\theta^i, \theta^{-i}) = - \int_0^{\theta^i} (s^i - x_2^i(s^i, \theta^{-i})) \frac{\partial x_2^i}{\partial s^i}(s^i, \theta^{-i}) ds^i + h^i(\theta^{-i}). \quad (5.3)$$

Una vez elegida la regla de x_2^i , la de x_1^i queda determinada por (5.3) salvo una función libre h^i de los tipos de los demás.

7.3 Condición de segundo orden: monotonicidad

Paso 1. Sea $\phi(\theta^i, \hat{\theta}^i) = \hat{\theta}^i x_2^i(\theta^i, \theta^{-i}) - \frac{1}{2} x_2^i(\theta^i, \theta^{-i})^2 + x_1^i(\theta^i, \theta^{-i})$ (primer argumento: lo declarado; segundo: el tipo verdadero). La identidad (5.2) dice $\frac{\partial \phi}{\partial \theta^i}(\theta^i, \theta^i) \equiv 0$.

Paso 2. Derivando esa identidad a lo largo de la diagonal (regla de la cadena, ambos argumentos se mueven),

$$\frac{d}{d\theta^i} \frac{\partial \phi}{\partial \theta^i}(\theta^i, \theta^i) = \frac{\partial^2 \phi}{(\partial \theta^i)^2} + \frac{\partial^2 \phi}{\partial \theta^i \partial \hat{\theta}^i} = 0 \implies \frac{\partial^2 \phi}{(\partial \theta^i)^2} = - \frac{\partial^2 \phi}{\partial \theta^i \partial \hat{\theta}^i}.$$

Paso 3. La condición de segundo orden del agente, $\frac{\partial^2 \phi}{(\partial \theta^i)^2} \leq 0$, equivale entonces a $\frac{\partial^2 \phi}{\partial \theta^i \partial \hat{\theta}^i} \geq 0$.

Paso 4. Como $\frac{\partial \phi}{\partial \hat{\theta}^i} = x_2^i$, derivando respecto de θ^i se obtiene $\frac{\partial^2 \phi}{\partial \theta^i \partial \hat{\theta}^i} = \frac{\partial x_2^i}{\partial \theta^i}$, de modo que

$$\frac{\partial x_2^i}{\partial \theta^i}(\theta^i, \theta^{-i}) \geq 0 \quad (x_2^i \text{ no decreciente en } \theta^i). \quad (\text{SOC})$$

7.4 Suficiencia y caracterización

Suficiencia. La diferencia de utilidad entre decir la verdad y declarar θ^i , usando (5.3), se reduce a

$$(\hat{\theta}^i - \theta^i) x_2^i(\theta^i, \theta^{-i}) \leq \int_{\theta^i}^{\hat{\theta}^i} x_2^i(s^i, \theta^{-i}) ds^i.$$

Si $\hat{\theta}^i > \theta^i$, como x_2^i es no decreciente, $x_2^i(\theta^i, \theta^{-i}) \leq x_2^i(s^i, \theta^{-i})$ en $[\theta^i, \hat{\theta}^i]$ e integrando se cumple; si $\hat{\theta}^i < \theta^i$ la desigualdad se invierte junto con el intervalo y vuelve a valer. Luego decir la verdad es óptimo. ■

Proposición 2 (Caracterización). *Un mecanismo diferenciable es a prueba de estrategias si y solo si $x_2^i(\cdot)$ es no decreciente en θ^i y x_1^i está dada por (5.3), salvo una función arbitraria $h^i(\theta^{-i})$.*

La condición de factibilidad dice que el consumo total no puede exceder los recursos totales, bien por bien. La detallamos en tres pasos.

Paso 1. (Factibilidad bien por bien.) Con bienes privados y dotaciones w_1^i, w_2^i , la suma de lo consumido de cada bien no puede superar la dotación agregada. Para el bien 2,

$$\sum_{i=1}^I x_2^i(\theta^i, \theta^{-i}) \leq \sum_{i=1}^I w_2^i.$$

Paso 2. (El bien 1 es el numerario.) Para el bien 1 la condición es $\sum_i x_1^i \leq \sum_i w_1^i$. Sustituyendo la expresión (5.3) de cada x_1^i ,

$$\sum_{i=1}^I h^i(\theta^{-i}) - \sum_{i=1}^I \int_0^{\theta^i} (s^i - x_2^i(s^i, \theta^{-i})) \frac{\partial x_2^i}{\partial s^i}(s^i, \theta^{-i}) ds^i \leq \sum_{i=1}^I w_1^i.$$

Paso 3. (El bien 2 pasa a ser público.) Si el bien 2 es público, hay un único nivel x_2 que todos consumen ($x_2^i = x_2$), producido a un costo $C(x_2)$ en unidades del bien 1. Ese costo se descuenta del numerario disponible, de modo que la factibilidad del bien 1 pasa a ser $\sum_i x_1^i + C(x_2) \leq \sum_i w_1^i$. Sustituyendo (5.3) se obtiene

$$C(x_2(\theta)) + \sum_{i=1}^I h^i(\theta^{-i}) - \sum_{i=1}^I \int_0^{\theta^i} (s^i - x_2(s^i, \theta^{-i})) \frac{\partial x_2}{\partial s^i}(s^i, \theta^{-i}) ds^i \leq \sum_{i=1}^I w_1^i. \quad (\text{F})$$

La tensión central aparece en (F): la única libertad que queda son las funciones h^i , y equilibrar el presupuesto exige ajustarlas de un modo que en general no es compatible con la optimalidad de Pareto.

8 Subasta de Vickrey y mecanismos de Clarke–Groves

8.1 La subasta de Vickrey

Ejemplo 2 (Segundo precio). Se vende un objeto indivisible a tres postores con valores privados $v_1 = 10, v_2 = 7, v_3 = 4$. Gana quien puja más alto y paga la segunda puja más alta. Aquí gana el postor 1 y paga 7, con ganancia $10 - 7 = 3$.

La verdad es dominante.

Paso 1. Si el postor 1 puja por encima de 7 (incluido su valor), gana y paga 7: su ganancia es 3, sin importar cuánto más pujan.

Paso 2. Si puja por debajo de 7, pierde el objeto y obtiene 0, renunciando a la ganancia de 3.

Paso 3. Un postor que pierde siendo veraz (el 3, de valor 4) solo ganaría pujando por encima de $7 > 4$, pagando más de lo que el objeto vale para él, con pérdida. ■

El pago del ganador es la externalidad que impone al resto: lo priva de un excedente igual a la segunda valoración. Generalizar esta idea da los mecanismos de Clarke–Groves.

8.2 El mecanismo de Clarke y la clase de Groves

Sea un proyecto binario $y \in \{0, 1\}$ con utilidad cuasilineal $u^i = x^i + v^i(y)$, $v^i(0) = 0$, $v^i(1) = \hat{\theta}^i$ (valor neto del proyecto para i). El óptimo de Pareto realiza el proyecto si $\sum_i \hat{\theta}^i \geq 0$.

Definición 7 (Clase de Groves; Clarke como caso pivotal). Decisión: $y(\theta) = 1 \iff \sum_i \theta^i \geq 0$.
Transferencia:

$$t^i(\theta) = \begin{cases} \sum_{j \neq i} \theta^j + h^i(\theta^{-i}), & \text{si } \sum_k \theta^k \geq 0, \\ h^i(\theta^{-i}), & \text{si } \sum_k \theta^k < 0, \end{cases}$$

con h^i arbitraria. El mecanismo de Clarke es el caso $h^i(\theta^{-i}) = -\max\{\sum_{j \neq i} \theta^j, 0\}$, que da $t^i = 0$ salvo cuando i es *pivotal*, y entonces cobra $|\sum_{j \neq i} \theta^j|$.

La verdad es una estrategia dominante. La idea es que la transferencia de Groves hace que la utilidad privada de cada agente coincida —salvo una constante que él no controla— con el bienestar social total. Como el mecanismo elige y para maximizar ese total, al agente le conviene que el mecanismo crea su valor verdadero. Lo verificamos.

Preliminares. Fijemos un agente i con valor verdadero $\hat{\theta}^i$. Sea θ^{-i} el perfil de reportes de los demás, que tomamos como dado (esto es lo que significa “estrategia dominante”: el argumento vale para cualquier θ^{-i}). Abreviamos

$$S = \sum_{j \neq i} \theta^j \quad (\text{la suma de los valores reportados por los demás}),$$

y notamos que tanto S como $h^i(\theta^{-i})$ *no dependen del reporte θ^i de i* : son constantes desde su punto de vista. La utilidad de i es cuasilineal, $u^i = \hat{\theta}^i y + t^i$, evaluada siempre en su valor *verdadero* $\hat{\theta}^i$ (mentir cambia lo que el mecanismo hace, no lo que i realmente valora).

Paso 1: la utilidad de i en cada decisión. Sustituyendo la transferencia de Groves,

$$u^i = \begin{cases} \hat{\theta}^i + S + h^i, & \text{si el mecanismo elige } y = 1, \\ h^i, & \text{si el mecanismo elige } y = 0. \end{cases}$$

Paso 2: qué decisión preferiría i . Restando los dos casos, i obtiene más con $y = 1$ que con $y = 0$ exactamente cuando

$$(\hat{\theta}^i + S + h^i) - h^i = \hat{\theta}^i + S \geq 0.$$

El término h^i se cancela: *no influye* en qué decisión prefiere i . Así, la decisión ideal para i es “ $y = 1$ si y solo si $\hat{\theta}^i + S \geq 0$ ”.

Paso 3: qué decisión toma el mecanismo. Por la regla eficiente, dado el reporte θ^i de i ,

$$\text{el mecanismo elige } y = 1 \iff \theta^i + \underbrace{\sum_{j \neq i} \theta^j}_{=S} \geq 0 \iff \theta^i + S \geq 0.$$

Es decir, el mecanismo aplica el mismo criterio del Paso 2, pero usando el valor *reportado* θ^i en lugar del *verdadero* $\hat{\theta}^i$.

Paso 4: por qué la verdad es óptima. Comparando los Pasos 2 y 3, basta con que el mecanismo aplique el umbral con el valor correcto $\hat{\theta}^i$. Eso se logra reportando $\theta^i = \hat{\theta}^i$: entonces “ $\theta^i + S \geq 0$ ” coincide con “ $\hat{\theta}^i + S \geq 0$ ”, y el mecanismo elige justamente la decisión que i prefiere.

Falta ver que mentir nunca ayuda. Una mentira solo importa si *cambia* la decisión respecto de la verdad, y solo puede cambiarla en estos dos sentidos, ambos perjudiciales para i :

- forzar $y = 1$ cuando la verdad daba $y = 0$, es decir cuando $\hat{\theta}^i + S < 0$: el agente recibe $\hat{\theta}^i + S < 0$ de más, peor que el 0 que tenía;
- forzar $y = 0$ cuando la verdad daba $y = 1$, es decir cuando $\hat{\theta}^i + S \geq 0$: el agente renuncia a un excedente $\hat{\theta}^i + S \geq 0$.

En ambos casos la mentira lo deja igual o peor, nunca mejor. Como esto vale para todo θ^{-i} y toda mentira θ^i , reportar la verdad es una estrategia dominante. ■

La razón de fondo. Tras la transferencia, $u^i = \hat{\theta}^i y + S + h^i = (\hat{\theta}^i + \sum_{j \neq i} \theta^j) y + h^i$; salvo la constante h^i , el agente maximiza la *suma de valores* —el bienestar social que el mecanismo ya optimiza—. Cada agente internaliza así el efecto de la decisión sobre los demás, y la honestidad deja de tener costo. Esta es exactamente la externalidad que el término $\sum_{j \neq i} \theta^j$ “le cobra” (o le paga) a i .

Proposición 3 (Green–Laffont). *La clase de Groves agota los mecanismos eficientes en estrategias dominantes; sin embargo, en general $\sum_i t^i(\theta) \leq 0$, de modo que el excedente recaudado no puede devolverse sin destruir los incentivos y la asignación no resulta óptima de Pareto.*

Ejemplo 3 (Clarke con tres agentes). Valores netos $\theta^1 = 5$, $\theta^2 = 4$, $\theta^3 = -7$. La suma es $5 + 4 - 7 = 2 \geq 0$, así que **el proyecto se realiza**.

La regla de Clarke cobra a cada agente *el daño que su presencia impone al resto*. Concretamente: i paga solo si es **pivotal**, es decir, si la decisión *con* i difiere de la que habrían tomado los demás *sin* él —decisión que depende del signo de $S_i := \sum_{j \neq i} \theta^j$, el valor neto del resto—. Cuando es pivotal, paga $|S_i|$; cuando no, paga 0.

agente	$S_i = \sum_{j \neq i} \theta^j$	sin i	con i	pivotal?	paga
1	$4 + (-7) = -3$	rechaza ($S_1 < 0$)	realiza	sí	$ -3 = 3$
2	$5 + (-7) = -2$	rechaza ($S_2 < 0$)	realiza	sí	$ -2 = 2$
3	$5 + 4 = 9$	realiza ($S_3 > 0$)	realiza	no	0

Lectura fila por fila:

- **Agente 1.** Sin él, los otros dos suman $-3 < 0$: rechazarían. Con él, el total es $+2$ y el proyecto *sí* se hace. Luego el agente 1 **voltea** la decisión; al hacerlo, fuerza un proyecto que el resto desvalora en 3. Clarke le cobra ese daño: 3.
- **Agente 2.** Mismo razonamiento: sin él, los otros suman $-2 < 0$ (rechazarían); con él se realiza. Es pivotal y paga el daño que causa al resto, 2.
- **Agente 3.** Tiene el valor más extremo (-7), pero *no* es pivotal: sin él los otros suman $9 > 0$ y el proyecto se haría igual. Su presencia no cambia nada, así que no impone daño y paga 0. *Pivotal no es lo mismo que “tener mucho valor”*: es *cambiar el resultado*.

8.3 Derivación diferencial para un bien público divisible

Con un bien público continuo y utilidad $x^i + v^i(y, \theta^i)$, la cantidad eficiente $y^*(\theta)$ queda definida por

$$\sum_{i=1}^I \frac{\partial v^i}{\partial y}(y, \theta^i) = 0. \quad (6.3)$$

Conviene leer (6.3) con cuidado. El término $\frac{\partial v^i}{\partial y}$ es la *valoración marginal* del agente i : cuántas unidades de bien privado (el numerario) gana por una unidad adicional de bien público. La condición (6.3) suma esas valoraciones marginales sobre todos los agentes y las iguala a cero; es la *condición de Samuelson* para este caso. Como la utilidad es $x^i + v^i$ sin un término de costo, producir bien público no cuesta numerario al margen, de modo que el nivel eficiente y^* es aquel donde el beneficio marginal social —la suma de las valoraciones marginales— se anula. Equivale a la condición de primer orden de maximizar el excedente total $\sum_i v^i(y, \theta^i)$ respecto de y : si $\sum_i \frac{\partial v^i}{\partial y} > 0$, entre todos valoran una unidad extra más de lo que cuesta y conviene subir y ; si es negativa, conviene bajarlo; en y^* se equilibran. (Con un costo $c(y)$ en numerario, el lado derecho sería $c'(y)$ en vez de 0.)

Paso 1. El agente i declara θ^i para maximizar $v^i(y^*(\theta^i, \theta^{-i}), \hat{\theta}^i) + t^i(\theta^i, \theta^{-i})$; su condición de primer orden es

$$\frac{\partial v^i}{\partial y} \frac{\partial y^*}{\partial \theta^i} + \frac{\partial t^i}{\partial \theta^i} = 0.$$

Paso 2. Volviéndola identidad en la diagonal se obtiene la ecuación diferencial

$$\frac{\partial t^i}{\partial \theta^i} = - \frac{\partial v^i}{\partial y}(y^*(\theta), \theta^i) \frac{\partial y^*}{\partial \theta^i}(\theta).$$

Paso 3. Integrando de 0 a θ^i ,

$$t^i(\theta) = - \int_0^{\theta^i} \frac{\partial v^i}{\partial y}(y^*(s^i, \theta^{-i}), s^i) \frac{\partial y^*}{\partial s^i}(s^i, \theta^{-i}) ds^i + \tilde{h}^i(\theta^{-i}). \quad (6.4)$$

Paso 4. De (6.3), $\frac{\partial v^i}{\partial y}(y^*(\theta), \theta^i) = - \sum_{j \neq i} \frac{\partial v^j}{\partial y}(y^*(\theta), \theta^j)$; sustituyendo en (6.4) y reorganizando,

$$t^i(\theta) = \sum_{j \neq i} v^j(y^*(\theta), \theta^j) + h^i(\theta^{-i}).$$

La transferencia de cada agente es la suma de las valoraciones de los demás (el valor social que genera para el resto) más una función libre h^i : es exactamente la clase de Groves, obtenida de forma constructiva.

9 Debilitar a Bayes–Nash: el mecanismo AGV

Ambiente bayesiano. Hay I agentes. El **tipo** de i es $\theta^i \in \Theta^i$ —su información privada, p. ej. cuánto valora el proyecto—. Los tipos se extraen de un **prior común** y, de manera crucial aquí, *independientes entre agentes*. La utilidad es cuasilineal,

$$u^i = v^i(y, \theta^i) + t^i,$$

donde v^i es el valor que i asigna a la decisión y y t^i su transferencia. La regla de decisión $y^*(\theta)$ es **eficiente**: maximiza $\sum_j v^j(y, \theta^j)$ en cada perfil.

Cuando i contempla su reporte, no conoce los tipos ajenos θ^{-i} : solo tiene su **creencia** sobre ellos, la densidad $\psi^i(\theta^{-i})$. Por la independencia, esta creencia *no depende de su propio tipo* θ^i . Promediar contra ψ^i es, por definición, tomar **esperanza** sobre los tipos ajenos:

$$\mathbb{E}_{\theta^{-i}}[g(\theta)] := \int_{\Theta^{-i}} g(\theta^i, \theta^{-i}) \psi^i(\theta^{-i}) d\theta^{-i}.$$

El concepto de solución se debilita: ya no pedimos veracidad pase lo que pase (dominantes), sino solo **en valor esperado** —equilibrio de **Bayes–Nash**—. Es decir, suponiendo que los demás reportan su verdad, a i le conviene reportar la suya *en promedio* sobre las creencias que tiene acerca de ellos. Se buscan reglas de pago $r^i(\theta^i)$ que induzcan esa veracidad y, además, equilibren el presupuesto.

Paso 1: el problema de i . Suponiendo veraces a los demás, i con tipo verdadero $\hat{\theta}^i$ elige su reporte θ^i para maximizar su utilidad **esperada**:

$$\max_{\theta^i} \mathbb{E}_{\theta^{-i}} \left[v^i(y^*(\theta^i, \theta^{-i}), \hat{\theta}^i) + r^i(\theta^i) \right] = \max_{\theta^i} \int_{\Theta^{-i}} \{v^i(y^*(\theta), \hat{\theta}^i) + r^i(\theta^i)\} \psi^i(\theta^{-i}) d\theta^{-i}.$$

(La esperanza recae solo sobre θ^{-i} ; el reporte θ^i entra a través de la decisión y^* .)

Paso 2: condición de primer orden. El reporte veraz $\theta^i = \hat{\theta}^i$ debe ser óptimo, así que la derivada respecto de θ^i se anula en él:

$$\mathbb{E}_{\theta^{-i}} \left[\frac{\partial v^i}{\partial y} \frac{\partial y^*}{\partial \theta^i} \right] + \frac{\partial r^i}{\partial \theta^i} = 0, \quad \iff \quad \int_{\Theta^{-i}} \frac{\partial v^i}{\partial y} \frac{\partial y^*}{\partial \theta^i} \psi^i(\theta^{-i}) d\theta^{-i} + \frac{\partial r^i}{\partial \theta^i} = 0.$$

El primer término es el efecto esperado de un reporte marginalmente mayor sobre el propio valor de i (vía el cambio en la decisión); r^i debe compensarlo.

Paso 3: integrando. Despejando $\partial r^i / \partial \theta^i$ e integrando desde un punto base 0,

$$r^i(\theta^i) = - \int_0^{\theta^i} \int_{\Theta^{-i}} \frac{\partial v^i}{\partial y} \frac{\partial y^*}{\partial s^i} \psi^i(\theta^{-i}) d\theta^{-i} ds^i + \text{cte.} \quad (7.1)$$

Paso 4: la externalidad esperada. Aquí interviene la identidad de Groves. Como y^* maximiza la suma total $\sum_j v^j$, en el óptimo el efecto marginal del propio valor de i se equilibra con el de los demás; sustituyendo en (7.1) e invirtiendo el orden de integración se obtiene

$$r^i(\theta^i) = \mathbb{E}_{\theta^{-i}} \left[\sum_{j \neq i} v^j(y^*(\theta), \theta^j) \right] + \text{cte} = \int_{\Theta^{-i}} \sum_{j \neq i} v^j(y^*(\theta), \theta^j) \psi^i(\theta^{-i}) d\theta^{-i} + \text{cte.} \quad (7.2)$$

La lectura es transparente: a i se le paga el **bienestar esperado de los demás** bajo la decisión eficiente —su *externalidad esperada* sobre el resto—. Como en Groves, esto alinea su interés con el total; la diferencia es que aquí la alineación es *en esperanza*, no estado por estado.

Paso 5: equilibrar el presupuesto. Las reglas (7.2) inducen veracidad pero no suman cero. El truco de d'Aspremont–Gérard-Varet es hacer que cada agente *financie* la externalidad esperada de los demás, repartida en partes iguales:

$$t^i(\theta) = r^i(\theta^i) - \frac{1}{I-1} \sum_{j \neq i} r^j(\theta^j).$$

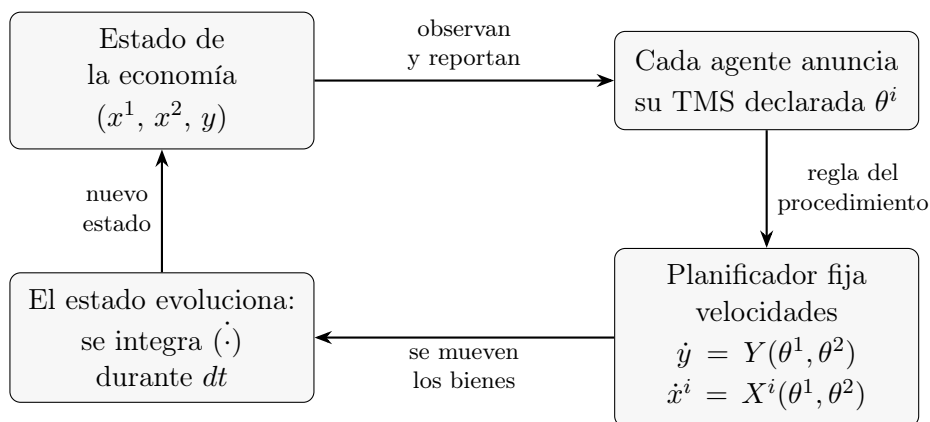
Restar $r^j(\theta^j)$ —que depende de θ^j , no de θ^i — no altera los incentivos de i (le es una constante en esperanza), pero la suma se cancela término a término:

$$\sum_{i=1}^I t^i(\theta) = \sum_i r^i(\theta^i) - \frac{1}{I-1} \sum_i \sum_{j \neq i} r^j(\theta^j) = \sum_i r^i(\theta^i) - \sum_j r^j(\theta^j) \equiv 0,$$

pues cada r^j aparece exactamente $I-1$ veces en el doble sumatorio.

10 Procedimientos de planificación

El planificador quiere fijar el nivel óptimo de un bien público, pero *no conoce cuánto lo valora cada quien*. En vez de preguntarlo de un golpe, monta un **proceso de tanteo** (*tâtonnement*): avanza el bien público de a poco, en cada instante pide a los agentes que declaren cuánto lo valoran *al margen*, y usa esas declaraciones para decidir hacia dónde y cuán rápido moverse. El reto de diseño es que a cada agente le *convenga decir la verdad* en cada instante. El ciclo es:



Los bienes (qué son x e y). Hay un **bien público** y , consumido por igual por los dos agentes (un parque, un dique: lo que uno disfruta no se lo quita al otro). Y hay un **bien privado** x^i para cada agente i —el numérario, “dinero”, divisible y transferible—. El **estado** de la economía en el instante t es la terna $(x^1(t), x^2(t), y(t))$, y la utilidad de i es $v^i(x^i(t), y(t))$, creciente en sus dos argumentos. Los superíndices 1, 2 son etiquetas de agente, no potencias.

El punto = velocidad. Para cualquier magnitud $z(t)$ escribimos

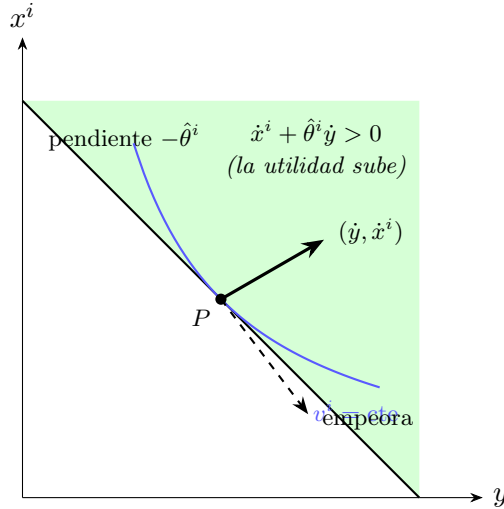
$$\dot{z} := \frac{dz}{dt} \quad (\text{su velocidad de cambio}).$$

Un *procedimiento* no fija niveles, sino **velocidades**: en cada instante decide \dot{y} (a qué ritmo crece el bien público) y \dot{x}^i (a qué ritmo se ajusta el privado de cada uno). El nivel se obtiene integrando esas velocidades en el tiempo.

La TMS, geoméricamente. La **tasa marginal de sustitución verdadera** de i ,

$$\hat{\theta}^i(x^i, y) := \frac{\partial v^i / \partial y}{\partial v^i / \partial x^i},$$

es cuántas unidades de bien privado está dispuesto i a ceder por una unidad marginal de bien público: su *disposición marginal a pagar*. En el plano (y, x^i) es, con signo, la **pendiente de su curva de indiferencia**: sobre una curva $v^i = \text{cte}$, $dx^i/dy = -\hat{\theta}^i$. Es información privada; el planificador no la ve.



Comportamiento miope. El agente no calcula la trayectoria entera; en cada instante mira solo la **tasa a la que mejora**, dv^i/dt . Por la regla de la cadena sobre $v^i(x^i(t), y(t))$, factorizando $\partial v^i/\partial x^i > 0$:

$$\frac{dv^i}{dt} = \frac{\partial v^i}{\partial x^i} \dot{x}^i + \frac{\partial v^i}{\partial y} \dot{y} = \underbrace{\frac{\partial v^i}{\partial x^i}}_{>0} (\hat{\theta}^i \dot{y} + \dot{x}^i).$$

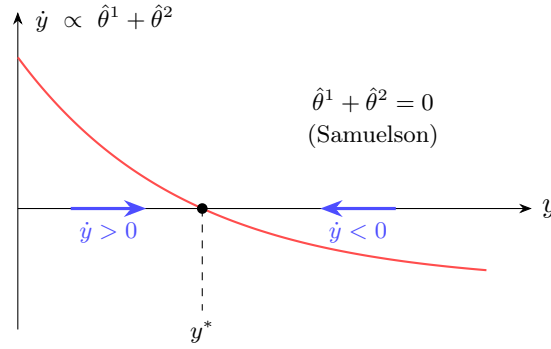
Como el factor de adelante es positivo, **el signo lo decide el paréntesis** $\hat{\theta}^i \dot{y} + \dot{x}^i$. Esto es justo lo que muestra el diagrama: el vector de movimiento (\dot{y}, \dot{x}^i) que el procedimiento imprime al estado mejora a i si y solo si apunta por encima de la tangente a su curva de indiferencia (región verde). El mismo avance del bien público $\dot{y} > 0$ puede beneficiarlo (flecha sólida) o perjudicarlo (flecha punteada) según cuánto privado se le quite a cambio: lo único que importa es de qué lado de la recta de pendiente $-\hat{\theta}^i$ cae.

El procedimiento directo. Como no observa $\hat{\theta}^i$, el planificador pide a cada agente que **anuncie** una tasa θ^i (su TMS declarada, quizá falsa) y, en función de los dos anuncios, fija

$$\dot{y} = Y(\theta^1, \theta^2), \quad \dot{x}^i = X^i(\theta^1, \theta^2), \quad i = 1, 2.$$

Por el **principio de revelación local**, no se pierde generalidad: si algún mecanismo indirecto induce veracidad instante a instante, hay uno directo equivalente. La pregunta de diseño es qué funciones Y, X^1, X^2 hacen que *anunciar la verdad* $\theta^i = \hat{\theta}^i$ sea óptimo para cada agente en cada instante; mirando el diagrama, que el vector (\dot{y}, \dot{x}^i) resultante caiga en la región verde de quien reporta su TMS real, gane lo que gane el otro. Esa es la propiedad **SLIIC**.

A dónde converge. Bajo veracidad, el ritmo del bien público responde a la suma de las TMS, $\dot{y} \propto \hat{\theta}^1(y) + \hat{\theta}^2(y)$. Como cada TMS decrece al saturarse el bien público, el sistema avanza mientras la suma sea positiva y se detiene donde se anula: esa es la **condición de Samuelson** $\sum_i \hat{\theta}^i = 0$. Es un único punto fijo, y es estable:



A la izquierda de y^* la suma de TMS es positiva ($\dot{y} > 0$, el bien público crece); a la derecha es negativa ($\dot{y} < 0$, se contrae). El proceso de tanteo lleva la economía a y^* desde cualquier inicio.

11 Resumen

Enfoque	Concepto	Eficiente / cuadra	Costo
General	dominante	imposible	dictadura (Gibbard–Satterthwaite)
Votante mediano	dominante	sí (sin transferencias)	preferencias unimodales
Clarke–Groves	dominante	decisión sí, presupuesto no	excedente que se quema
AGV	Bayes–Nash	sí (presupuesto cuadra)	prior común conocido
MDP	maximin local	converge a Pareto	proceso dinámico, miopía

Cuadro 1: Compromiso entre veracidad robusta y eficiencia plena.

El método común es el enfoque diferencial: convierte “decir la verdad es óptimo” en una condición de primer orden que, vuelta identidad e integrada, fija la forma del mecanismo ((5.3), (6.4), (7.1)–(7.2), (8.2)); la condición de segundo orden se reduce a una monotonicidad. El conflicto central: con cuasilinealidad se obtiene veracidad robusta (Groves, presupuesto que sobra) o eficiencia plena (AGV, presupuesto que cuadra), pero en general no ambas.

Referencias

- Gibbard, A. (1973). *Manipulation of voting schemes*. *Econometrica* 41, 587–601.
- Satterthwaite, M. (1975). *Strategy-proofness and Arrow’s conditions*. *J. Econ. Theory* 10, 187–217.
- Mount, K. & Reiter, S. (1974). *The informational size of message spaces*. *J. Econ. Theory* 8, 161–192.
- Clarke, E. (1971). *Multipart pricing of public goods*. *Public Choice* 8, 19–33.

- Groves, T. & Ledyard, J. (1977). *Optimal allocation of public goods*. *Econometrica* 45, 783–810.
- Vickrey, W. (1961). *Counterspeculation, auctions and competitive sealed tenders*. *J. Finance* 16, 1–17.
- d'Aspremont, C. & Gérard-Varet, L.A. (1979). *Incentives and incomplete information*. *J. Public Econ.* 11, 25–45.
- Dreze, J. & de la Vallée Poussin, D. (1971). *A tâtonnement process for public goods*. *Rev. Econ. Studies* 38, 133–150.
- Green, J. & Laffont, J.J. (1979). *Incentives in Public Decision Making*. North-Holland.