

## Más Allá de Pagar por el Desempeño

Bengt Holmström

Massachusetts Institute of Technology, Cambridge, MA, USA. Deseo agradecer a George Baker, Robert Gibbons, Oliver Hart, Paul Milgrom, Canice Prendergast, John Roberts y Jean Tirole por años de fructíferas discusiones sobre el tópico de esta conferencia y a Jonathan Day, Dale Delitis, Robert Gibbons, Gary Gorton, Parag Pathak, Alp Simsek, David Warsh y en especial a Iván Werning [actual asesor del BCRA] por sus comentarios sobre distintas versiones de este paper.

[Pay For Performance and Beyond](#)  
*Nobel Prize Lecture, December 8, 2016*

*Traducción de Enrique A. Bour*

En esta conferencia, hablaré sobre mi trabajo en contratos de incentivo, especialmente de incentivos relacionados con riesgo moral. Proporcionaré una narración de mi recorrido intelectual desde modelos estrechamente enfocados en el pago por desempeño, hasta modelos que ven el alcance del problema de incentivos en términos mucho más amplios, caracterizando a empleados y empresas multitareas que hacen uso extensivo de instrumentos no financieros para diseñar sistemas de incentivos coherentes.

Resaltaré los momentos clave de este viaje, incluyendo los malentendidos y los nuevos puntos de vista. Los primeros a menudo son precursores de estos últimos. En el proceso, espero transmitir una idea de cómo trabajo con modelos. No hay una forma correcta de teorizar, pero creo que es importante desarrollar un estilo consistente con el que uno se sienta cómodo.



[Bengt Holmström](#) Finlandés, n. en 1949 Nobel 2016  
[Pay for Performance and Beyond](#) 36m  
[Hölmstrom & Milgrom, The Firm as an Incentive System](#) 1994

Comenzaré con una breve reseña de la forma indirecta en que me convertí en economista. Esto revelará los orígenes de mi interés en los problemas de incentivos y dará cuenta de mi asociación de toda la vida con la práctica empresarial, lo que influyó fuertemente sobre mi investigación y estilo de trabajo.

No planeaba convertirme en un académico. Después de graduarme en la Universidad de Helsinki, conseguí un trabajo en Ahlstrom como planificador corporativo. Ahlstrom era una de las diez empresas más grandes de Finlandia en ese momento, un gran conglomerado con 20-30 fábricas en todo el país. La compañía se enorgullecía de estar al tanto de los avances más recientes en administración. Me contrataron para implementar un modelo de programación lineal que se había elaborado recientemente para ayudar a la administración con la planificación estratégica a largo plazo. Era un modelo enorme con miles de variables y cientos de restricciones, que describen todas las fábricas del conglomerado, sus actividades de producción, planes de inversión e interdependencias financieras y tecnológicas. La compañía había invertido mucho en computadoras mainframe y tenía grandes esperanzas de que el modelo de planificación fuera uno de los factores decisivos de sus inversiones.

Mi tarea más urgente fue organizar el proceso de recopilación de datos. Comencé a visitar fábricas para analizar los datos que se requerían. No tardé mucho en apreciar que las personas que proporcionaban los datos desconfiaban profundamente de un matemático de 23 años enviado desde la sede central para recopilar datos para un modelo de planificación que aconsejaría a la alta dirección sobre cómo administrar los escasos recursos en inversiones. Querían saber con qué números alimentar mi modelo de "caja negra" para garantizar que sus propios planes para su fábrica recibieran la cantidad apropiada de recursos. Los datos fueron llegando en forma lenta y vacilante.

Después de algunos meses llegué a la conclusión de que toda la tarea estaba mal orientada. Aun suponiendo que hubiera las mejores intenciones a nivel de fábrica, sería muy difícil obtener datos razonables. La calidad de los datos variaba mucho y a menudo surgían desacuerdos, sobre todo en casos en que las fábricas estaban interconectadas. Además, cada vez me preocupaba más el carácter estratégico de los datos proporcionados. En consecuencia la seriedad de los datos parecía cuestionable tanto por razones técnicas como estratégicas.

Sugerí que renunciáramos al gran proyecto y que en cambio me enfocara en dos cosas: (i) modelos más pequeños que podrían ayudar a cada fábrica a mejorar su propio proceso de planificación, y (ii) tratar de lidiar con los problemas de incentivos en la planificación de inversiones a nivel corporativo.

Mi primera recomendación tuvo cierto éxito. Desarrollé pequeños programas lineales que parecían susceptibles de ser aplicados a las fábricas. Utilicé estos modelos como lo haría un economista: intenté replicar lo que las fábricas estaban haciendo. Esto significaba un montón de ida y vuelta. Conseguía los datos, ejecutaba el programa lineal y luego le decía a la fábrica lo que el programa estaba proponiendo como "solución óptima" dados sus datos y, sobre todo, por qué el modelo proponía la solución que generaba. Este último paso, el que explica cómo estaba pensando el modelo, fue clave. Dejaba en claro que yo no estaba allí para recomendar una solución mecánica; estaba allí para entender lo que podía faltar en la especificación de mi modelo. Esta lección -la utilidad de los modelos pequeños y de escucharlos - estaría junto a mí a lo largo de mi carrera académica.

Mi segunda recomendación, pensar creativamente en los problemas de incentivos que rodean la planificación de inversiones, fue un fracaso. Cometí todos los errores que uno probablemente cometa cuando se trata de diseñar incentivos por primera vez. Mi pensamiento se guiaba por dos principios: las fábricas deberían pagar sus préstamos y el precio debería surgir, al menos en parte, a través de un proceso de mercado, de modo que los fondos se asignaran de forma eficiente. En lenguaje de hoy en día, sugerí que "lleváramos el mercado al interior de la empresa" para asignar fondos a las fábricas.

Hoy, lo sé mejor. Como trataré de explicar, una de las principales lecciones de trabajar en problemas de incentivos durante 25 años es que dentro de las empresas, los incentivos financieros de alta potencia pueden ser muy disfuncionales y los intentos de llevar el mercado al interior de la empresa generalmente son un mal consejo. Por lo general, es mejor evitar los incentivos de alta potencia y, en ocasiones, no utilizar el pago por desempeño en absoluto. El reciente escándalo en Wells Fargo explica el motivo (Tayan 2016). Los incentivos monetarios, potentes pero desalineados, llevaron a algunos gerentes a vender cuentas falsas para mejorar sus bonos. El famoso artículo de Kerr (1975) [\*The Folly of Hoping for A, While Paying for B\*](#) podría haber servido como advertencia [*ver cuadro pág. sgte.*], aunque el artículo es bastante ligero sobre sugerencias para proporcionar formas alternativas de incentivos. Espero mostrar que nuestra comprensión de los problemas de incentivos ha avanzado bastante desde los días en que Kerr publicó su artículo. Para apreciar el progreso del pensamiento, comenzaré con la literatura inicial sobre modelos de principal-agente.

**The Folly of Rewarding A,  
While Hoping for B**

□ Involves rewarding behavior that is trying to be discouraged, while the desired behavior is not being rewarded at all

MANAGERS HOPE FOR:	BUT MANAGERS FREQUENTLY REWARD:
Long-term growth and environmental/social responsibility	Quarterly earnings
Innovative thinking and risk-taking	Proven methods and not making mistakes
Teamwork and collaboration	The best competitive individual performers
Employee involvement and empowerment	Tight control over operations and resources
High achievement	Another year's effort
Candor such as telling of bad news early	Reporting good news, whether it is true or not, and agreeing with the boss, whether the boss is right or wrong

Source: S. Bett, "On the Folly of Rewarding A, While Hoping for B," *Journal of Management Education* 9 (February 1995): 30-41.

## I. El problema Principal-Agente

### A. El Modelo de "Esfuerzo" Unidimensional

Las primeras contribuciones a la literatura principal-agente sobre riesgo moral incluyen Wilson (1969), Spence y Zeckhauser (1971), Ross (1973), Stiglitz (1975) y Mirrlees ([1975] 1999). Wilson y Ross se preguntaron bajo qué condiciones las preferencias del principal y del agente sobre loterías riesgosas estarán perfectamente alineadas compartiendo el riesgo en forma óptima. Esto es posible si las funciones de utilidad del principal y del agente son tales que el reparto lineal de riesgos sea óptimo. Spence y Zeckhauser (1971) estudiaron los contratos de seguro bajo distintos supuestos de información, incluido el caso de riesgo moral, así como el de selección adversa.

Mirrlees fue el primero en estudiar en general el caso en que el agente presta un servicio al principal y el problema es cómo motivar al agente para que trabaje diligentemente. La acción del agente a menudo es denominada "*esfuerzo*", aunque esta interpretación no debe ser tomada literalmente. El modelo es aplicable en una amplia gama de situaciones: un empleado que trabaja para un gerente; un abogado que sirve a un cliente; un doctor tratando a un paciente; o un CEO que sirve al consejo de una compañía, por nombrar algunos.

Hay dos desafíos al diseñar un esquema de incentivos óptimo para el esfuerzo. En primer lugar, el agente considera que es privadamente costoso proporcionar el servicio, al menos más allá de cierto nivel básico, por lo que se necesita un incentivo financiero basado en el rendimiento. Pero el rendimiento es imperfectamente medido, por lo que la remuneración variable inducirá riesgos para el agente. Como el agente es intolerante al riesgo, existe una relación de compromiso entre riesgo e incentivo. ¿Cómo debería resolverse de manera óptima?

Se podría abordar este problema estudiando esquemas simples de incentivos, como un incentivo lineal además de un salario fijo, o una bonificación por el rendimiento más allá de un estándar mínimo.<sup>1</sup> El problema con el uso de una forma funcional particular es que

<sup>1</sup> Es común ver a las empresas pagando a los ejecutivos una bonificación que es lineal dentro de un intervalo de rendimiento, pero limitado tanto en la parte superior como en la inferior; ver Murphy (1999). Por otro lado, los agentes y vendedores de bienes raíces reciben comisiones pagadas sin un

el análisis no nos dirá por qué se usan diferentes incentivos en diferentes contextos. Además, fijar la forma de la paga de incentivo puede silenciar compensaciones que son esenciales para comprender el problema de incentivo subyacente. Esto hace que sea valioso estudiar el problema sin restricciones funcionales.

Permítaseme pasar a una formulación simple y genérica de la relación principal-agente. El agente elige un nivel de esfuerzo no observado  $e$ . La elección de esfuerzo del agente conduce a un pago  $x = x(e, \varepsilon)$ , donde  $\varepsilon$  capta factores externos aleatorios, como condiciones de mercado o errores de medición que el agente no controla directamente. A menudo trabajaré con la especificación aditiva  $x = e + \varepsilon$ .

Como señaló Mirrlees ([1975] 1999), es técnicamente conveniente y más elegante ver al agente eligiendo una distribución sobre  $x$ . Para una elección fija  $e$ , la distribución sobre  $\varepsilon$  induce una distribución sobre  $x$ , denotada  $F(x|e)$ . Esto elimina  $\varepsilon$  y proporciona una caracterización más simple y, como veremos, mucho más informativa, aunque la dependencia explícita de  $e$  puede ser útil para analizar ciertos contextos.<sup>2</sup>

Antes de que actúe el agente, el principal le ofrece al agente un contrato de incentivo, que le paga al agente  $s(x)$  cuando el resultado es  $x$ . El principal retiene la diferencia  $x - s(x)$ . Las utilidades del agente y del principal son, respectivamente,  $U = u(s(x)) - c(e)$  y  $V = x - s(x)$ , por lo que el principal es neutro al riesgo y el agente (en general) es intolerante al riesgo. La función de utilidad del agente es aditivamente separable, lo que es restrictivo pero es comúnmente usado.

El principal y el agente están informados simétricamente en el momento de firmar el contrato (esto es lo que hace que el problema sea de *riesgo moral*). En particular, conocen las funciones de utilidad del otro y tienen las mismas creencias sobre las distribuciones  $F(x|e)$ . Por lo tanto, el principal puede pronosticar el comportamiento del agente dado  $s(x)$  aunque no pueda observar la elección de su esfuerzo.

El esquema de incentivos  $s(x)$  debe proporcionar al agente una utilidad esperada que sea por lo menos tan elevada como la que el agente puede obtener en otro lado. La participación del agente requiere que el director considere el impacto  $s(x)$  sobre la utilidad esperada del agente. La carga del agente por el riesgo extra y el esfuerzo adicional corre a cargo del principal. Hallar el mejor contrato es, por lo tanto, un interés compartido en el modelo (pero no necesariamente en la práctica).

Para determinar la oferta óptima del principal, es útil pensar que éste propone un nivel de esfuerzo  $e$  junto con un esquema de incentivos  $s(x)$  tales que el agente esté conforme de elegir  $e$ , es decir, que  $s(x)$  y  $e$  son incentivo-compatibles. Esto lleva al siguiente programa para encontrar el par óptimo  $\{s(x), e\}$ :

- (1)  $\text{Max } E[x - s(x) | e]$ , sujeto a
- (2)  $E[u(s(x)) - c(e) | e] \geq E[u(s(x)) - c(e') | e']$  para  $e' \neq e$ , y
- (3)  $E[u(s(x)) - c(e) | e] \geq U_r$ .

La primera restricción asegura que  $e$  sea óptimo para el agente. La segunda restricción

---

límite superior. El artículo de Stiglitz (1975) sobre aparcería fue el primero en estudiar incentivos lineales.

<sup>2</sup> Mirrlees ([1975] 1999) fue el primero en usar esta formulación. Evita tomar derivadas del esquema de incentivos endógeno  $s(x)$ , que bien puede ser a priori no diferenciable.

garantiza que el agente obtenga al menos su utilidad de reserva  $U_r$  si elige  $e$  y, por lo tanto, aceptará el contrato. Supondré que existe una solución óptima para este programa.<sup>3</sup>

### B. Casos de óptimo primario

Antes de pasar a analizar la solución óptima para el programa de óptimo secundario (1) - (3) es útil analizar algunos casos donde la solución óptima coincide con la solución de óptimo primario que se obtiene cuando la restricción de incentivo (2) puede ser eliminada, porque todo nivel de esfuerzo  $e$  puede aplicarse sin costo. Debido a que el principal es neutro al riesgo, el esfuerzo en el óptimo primario, denotado  $e^{OP}$ , maximiza  $E(x|e) - c(e)$ .

Una solución de óptimo primario se puede lograr en tres casos:

- I. No hay incertidumbre.
- II. El agente es neutro al riesgo.
- III. La distribución tiene soporte móvil.

Si es válido (I), el agente elegirá  $e^{OP}$  si recibe el salario fijo  $w = u^{-1}(U_r + c(e^{OP}))$  siempre que  $x \geq e^{OP}$  y nada en caso contrario. El salario  $w$  será suficiente para coincidir con la utilidad  $U_r$  de reserva del agente. Si se cumple (II), el principal puede rentar la tecnología al agente fijando  $s(x) = x - E(x | e^{OP}) + w$ . Como demandante residual neutro al riesgo, el agente elegirá  $e^{OP}$  y ganará su utilidad de reserva  $U_r$  como en el caso (I).

El tercer caso es el más interesante y también insinúa la forma en que razona el modelo. Para concretar, supóngase que  $x = e + \varepsilon$  siendo  $\varepsilon$  con una distribución uniforme en  $[0, 1]$ . Esto corresponde al agente que elige cualquier distribución uniforme  $[e, e + 1]$  al costo  $c(e)$ . La densidad de una distribución uniforme se parece a una caja. A medida que varía  $e$ , la caja se mueve hacia la derecha. En primer lugar, el agente debe recibir un pago constante si elige el mejor nivel de esfuerzo  $e^{OP}$ . Esto se puede implementar pagando al agente un salario fijo si el resultado observado es  $x \geq e^{OP}$  y algo lo suficientemente bajo (un castigo) si es  $x < e^{OP}$ . El esquema funciona, porque se dan dos condiciones: (a) el agente puede estar seguro de evitar el castigo eligiendo el mejor nivel de esfuerzo y (b) el soporte móvil le permite al principal inferir con certeza que un agente está holgazaneando si  $x < e^{OP}$  y, por lo tanto, castigarlo lo suficiente como para que elija el óptimo primario. En el modelo general, las inferencias son siempre imperfectas, pero seguirán desempeñando un papel central en la compensación del riesgo versus los incentivos.

### C. Óptimo Secundario con Dos Acciones

Procuró ahora caracterizar el esquema de incentivos óptimo en el caso especial donde el agente elige entre sólo dos distribuciones  $F_L$  y  $F_H$ . Este caso especial revelará la mayoría de los conocimientos del modelo básico de agencia sin tener que lidiar con complicaciones técnicas. Supongamos que  $F_H$  domina la distribución  $F_L$  en el sentido de dominancia estocástica de primer orden: para cualquier  $z$ , la probabilidad de que  $x > z$  es mayor en  $F_H$  que en  $F_L$ . Esto es consistente con suponer que la distribución alta es una opción más costosa para el agente:  $c_L < c_H$ . Por ejemplo, para  $x = e + \varepsilon$  y  $e_L < e_H$ ,  $F_H$  en primer orden domina estocásticamente  $F_L$  independientemente de cómo se distribuye.

<sup>3</sup> Holmström (1977) prueba la existencia suponiendo que  $s(x)$  debe elegirse de un intervalo finito. Una prueba de existencia más instructiva es provista por Grossman y Hart (1983) cuando el número de resultados es finito. La hipótesis clave es que las probabilidades están inferiormente acotadas por cero para todas las opciones.

Supóngase que el principal quiere implementar  $H$ , pues el otro caso no es interesante ya que  $L$  se implementa de manera óptima con un pago fijo. Sean  $\mu$  y  $\lambda$  los multiplicadores lagrangianos (no negativos) asociados con la restricción de compatibilidad de incentivos (2) y la restricción de participación (3) en el programa (1) - (3) del principal. El mejor contrato de óptimo secundario, denotado  $s_H(x)$ , se caracteriza por

$$(4) \quad u'(s_H(x))^{-1} = \lambda + \mu [1 - f_L(x)/f_H(x)], \text{ para todo } x. ^4$$

Aquí  $f_L(x)$  y  $f_H(x)$  son las funciones de densidad de  $F_L(x)$  y  $F_H(x)$ . Es fácil ver que ambas restricciones (2) y (3) son limitantes y, por lo tanto,  $\lambda$  y  $\mu$  son estrictamente positivas.<sup>5</sup>

La caracterización es simple pero informativa. En primer lugar, téngase en cuenta que el esquema de incentivos óptimo se desvía del óptimo primario, que le paga al agente un salario fijo, porque el lado derecho varía con  $x$ . La razón, por supuesto, es que el principal debe proporcionar un incentivo para que el agente ejerza un gran esfuerzo. En segundo lugar, la forma del esquema de incentivos óptimo sólo depende de la relación  $f_H(x)/f_L(x)$ . En estadística, esta relación es conocida como **razón de verosimilitud**; denótese como  $l(x)$ . La razón de verosimilitud en  $x$  indica cuán probable es que el resultado observado  $x$  se origine en la distribución  $H$  en lugar de la distribución  $L$ . Un valor superior a 1 habla a favor de  $H$  y un valor inferior a 1 habla en favor de  $L$ .

[La razón de verosimilitud (o *likelihood ratio*) es de uso amplio en diversas disciplinas, por ejemplo en medicina. Véase v.g. C. Silva Fuente-Alba y M. Molina Villagra, [Likelihood ratio \(razón de verosimilitud\): definición y aplicación en Radiología](#), *Rev. Argentina de Radiología*, 2017; 81(3): 204-208. La  $l(x)$  es definida como la razón entre la posibilidad de observar un resultado en los pacientes con la enfermedad en cuestión versus la posibilidad de ese resultado en pacientes sin la patología; la aplicabilidad de la  $l(x)$  es múltiple en la práctica radiológica, ya que es posible utilizarla al analizar pruebas diagnósticas con resultados *dicotómicos*, en los que sólo es posible determinar presencia o ausencia de enfermedad (negativo o positivo), o bien con resultados *categoricos*, por ejemplo mediante exámenes que tienen diferentes categorías de severidad (leve, moderada o severa). (N. del T.)]

Denótese por  $s_\lambda$  el valor constante de  $s(x)$  que satisface (4) con  $\mu = 0$ . Es el contrato óptimo de distribución de riesgos correspondiente a  $\lambda$ .<sup>6</sup> El contrato de óptimo secundario ( $\mu > 0$ ) se desvía de la distribución óptima del riesgo (pago fijo  $s_\lambda$ ) de una manera muy intuitiva. El agente es castigado cuando  $l(x)$  es menor que 1, porque  $x$  es evidencia de alto esfuerzo. Al agente se le paga una bonificación cuando  $l(x)$  es mayor que 1 porque la evidencia está a favor de un gran esfuerzo. Los desvíos son más grandes cuanto más fuerte sea la evidencia. Por lo tanto, el esquema de óptimo secundario está diseñado como si el principal estuviera extrayendo inferencias sobre la elección del agente, como en estadística. Esto es bastante sorprendente porque en el modelo el principal sabe que el agente está eligiendo un alto esfuerzo dado el contrato que ofrece al agente antes de que se observe el resultado  $x$ . Luego, no hay nada que inferir en el momento en que tiene lugar el resultado.

## II. El Principio de Informatividad

**El hecho de que el modelo de agencia básica piense como un estadístico es muy útil para comprender su comportamiento y predicciones.** Un caso importante es la respuesta que el

<sup>4</sup> Esta ecuación es la condición de primer orden del lagrangiano con respecto a  $s(x)$  para todo  $x$ .

<sup>5</sup> Si  $\mu = 0$ , la fórmula implica que  $s_H(x)$  es constante en cuyo caso el agente elegiría  $L$ , violando la restricción de incentivo; luego  $\mu > 0$ . Si  $\lambda = 0$  la restricción (2) no sería limitante y el principal podría hacerlo mejor reduciendo todos los niveles de utilidad en una constante; luego  $\lambda > 0$ .

<sup>6</sup> Por lo general,  $s_\lambda$  no corresponde a una distribución de riesgos óptima para el valor de reserva de  $U_T$  en la restricción de participación del agente (3), porque ese problema tendrá un valor de  $\lambda$  diferente. Le agradezco a Jörgen Weibull por señalar esto.

modelo da a la pregunta: ¿cuándo será valiosa una señal adicional, dado que le permite al principal redactar un mejor contrato?

A. Señales Adicionales

Podría pensarse que si  $y$  es suficientemente ruidoso, esto podría saturar el valor de cualquier información adicional incluida en  $y$ . Esta intuición es incorrecta. Esto se ve fácilmente con una extensión menor de (4). Hacemos que el esquema de incentivos óptimo que implementa  $H$  utilizando ambas señales sea  $s_H(x, y)$ . La caracterización de este esquema sigue exactamente los mismos pasos que para  $s_H(x)$ . El único cambio que hay que hacer en (4) es escribir  $s_H(x, y)$  en lugar de  $s_H(x)$  y la densidad conjunta  $f_i(x, y)$  en lugar de  $f_i(x)$  para  $i = L, H$ . Por supuesto, los multiplicadores de Lagrange no tendrán los mismos valores si  $y$  es valioso.

Considerando esta variante (4) se aprecia que si la razón de verosimilitud  $l(x, y) = f_H(x,y)/f_L(x, y)$  depende tanto de  $x$  como de  $y$ , luego en el primer miembro la solución óptima  $s_H(x, y)$  debe depender también de  $x$  y de  $y$ . En tal caso  $y$  es valioso. Por el contrario, si  $l(x, y)$  no depende de  $y$ , el esquema óptimo es de la forma  $s_H(x)$ . Podemos en tal caso escribir

(5)  $f(x, y|e) = g(x|e)h(y|x)$ ,  
para todo  $x, y$  y  $e = L, H$ .

**Informatividad**

Evalúa hasta qué punto las secuencias de un texto son predecibles o inesperadas, si transmiten información conocida o novedosa. Todo texto es en alguna medida informativo, esto es, aporta información nueva. Un alto nivel de informatividad puede requerir un esfuerzo mayor de procesamiento. Un bajo nivel de informatividad puede provocar fastidio o incluso rechazo.

*El concepto de Informatividad es usado habitualmente en Análisis Textual*

A la izquierda está la función de densidad para la distribución conjunta de  $x$  e  $y$  dada  $e$ . A la derecha, el primer término es la densidad condicional de  $x$  dado  $e$  y el segundo término la densidad condicional de  $y$  dado  $x$ . La clave es que la densidad condicional de  $y$  no depende de  $e$  y, por lo tanto,  $y$  no contiene ninguna información adicional sobre  $e$  (dada  $x$ ). En palabras tenemos lo siguiente:

**Principio de Informatividad** (Holmström, 1979, Shavell, 1979). - Una señal adicional  $y$  es valiosa si y sólo si aporta información adicional sobre lo que hizo el agente dada la señal  $x$ .

Dicho de esta manera, el resultado parece bastante obvio, pero sólo subraya el valor de usar la formulación de la función de distribución.<sup>7</sup> Se puede derivar una caracterización similar a (4) usando la formulación de espacio de estado  $x(e, \epsilon)$ , pero esta caracterización es difícil de interpretar porque no admite una interpretación estadística. La razón es que hay dos formas en que  $y$  puede ser informativo sobre  $e$  dado  $x$ . Podría ser que  $y$  proporcione otra señal directa sobre  $e$  ( $y = e + \delta$ , donde  $\delta$  es el ruido) o  $y$  proporcione informa-

<sup>7</sup> El principio de informatividad tal como se estableció se aplica cuando sólo hay una restricción vinculante de compatibilidad de incentivos (2). Si el agente está indiferente entre varias acciones, habrá un multiplicador de Lagrange  $\mu_j$  para cada restricción de incentivo vinculante  $j$ -ésima. Se puede extender la interpretación inferencial a esta situación evaluando el resultado frente a la probabilidad de que el agente randomice distintas acciones de indiferencia con los multiplicadores de Lagrange que proporcionan los pesos (relativos) de esta estrategia aleatorizada. (Agradezco a Paul Milgrom por sugerir esta extensión.) Gjesdal (1982) y Grossman y Hart (1983) proporcionan un ordenamiento más débil utilizando el teorema de Blackwell, que se aplica independientemente del número de restricciones de incentivo vinculantes. Para otras variantes, ver Kim (1995) y Chaigneau, Edmans y Gottlieb (2014).

ción acerca de  $\varepsilon$  ( $y = \delta$ , donde  $\delta$  y  $\varepsilon$  están correlacionados), lo que es información indirecta sobre  $e$ . Ambos canales son captados por el criterio único de informatividad.

Como ilustración del principio de informatividad, considérese el uso de **deducibles** en el siguiente entorno de seguro. Un asegurado puede tomar una medida para prevenir un accidente (un robo, por ejemplo). Pero dado que el accidente ocurre, la cantidad de daño causado no depende de la medida preventiva tomada por el asegurado. En este caso, es óptimo que el asegurado pague un deducible si el accidente sucede como incentivo para tomar precauciones, pero la compañía de seguros debe pagar todos los daños, independientemente de la cantidad. Esto es así porque la ocurrencia del accidente contiene toda la información relevante sobre las precauciones que tomó el agente. La cantidad de daño no agrega ninguna información sobre precauciones.

### B. Implicancias del Principio de Informatividad

Varias consecuencias se derivan directamente del principio de informatividad.

1. La aleatorización es subóptima. Condicional a la señal  $x$ , no sería deseable aleatorizar el pago al agente, ya que eliminando la aleatoriedad sin alterar la utilidad del agente en  $x$  le proporciona al principal una mayor rentabilidad. La aleatorización puede resultar óptima si la función de utilidad no es separable.<sup>8</sup>
2. La evaluación relativa del desempeño (Baiman y Demski, 1980, Holmström, 1982) es valiosa cuando el desempeño de otros agentes informa algo acerca de los factores externos que afectan el desempeño del agente, ya que la información relacionada con el desempeño medido  $x$  es informativa sobre la acción del agente  $e$ . La importante literatura de torneos iniciada por Lazear y Rosen (1981) estudia la evaluación relativa del rendimiento en profundidad.
3. El principio de controlabilidad en contabilidad establece que el incentivo de un agente sólo debe depender de los factores que el agente pueda controlar. Usar la evaluación relativa del desempeño parece violar este principio, ya que el agente no controla lo que hacen otros agentes. La interpretación correcta del principio de controlabilidad dice que un agente debe recibir un pago basado en la medida de rendimiento más informativa disponible. Como  $x$  ya depende de factores externos (capturados por  $\varepsilon$ ), cualquier cosa correlacionada con  $\varepsilon$  puede ser usada para filtrar el riesgo externo, haciendo que la medida de rendimiento ajustado sea más informativa.
4. Un estadístico suficiente en sentido estadístico también es suficiente para diseñar contratos de incentivos óptimos. Por ejemplo, la media muestral obtenida de forma independiente de una distribución normal con varianza conocida es un estadístico suficiente del esfuerzo medio y se puede usar para evaluar el rendimiento (Holmström 1982).
5. El pago de incentivo óptimo dependerá de información retrasada si la información valiosa aparece con retraso.

Bebchuk y Fried (2004), entre otros, han argumentado que no se debe permitir que los CEOs disfruten de ganancias inesperadas derivadas de condiciones macroeconómicas favorables. Apelando al principio de informatividad, defienden el uso de la evaluación relativa del desempeño como una forma de **filtrar la "suerte"**. En muchos casos esto está justificado, pero no sin calificaciones. En un contexto multitarea, la evaluación relativa del desempeño puede distorsionar la asignación de tiempo y esfuerzo del agente. Cuando los

<sup>8</sup> Véase Gjesdal (1982).



agentes que trabajan en conjunto son comparados entre sí, se ve perjudicada la cooperación o puede producirse colusión (Lazear 1989). En general, no es aconsejable filtrar los efectos de las variaciones en los precios del petróleo sobre la remuneración de los ejecutivos del petróleo (al menos en su totalidad) porque esto distorsionaría la inversión y otras decisiones críticas.

### C. *Enigmas y Deficiencias*

El principio de informatividad capta la lógica central del modelo básico de esfuerzo unidimensional. De este modo, ayuda a explicar algunas características desconcertantes del modelo básico y también sus principales deficiencias.

*Sorprendentemente, el esquema de incentivos óptimo no requiere ser no decreciente cuando  $x = e + \varepsilon$ .* La razón es que una producción más alta puede no estar señalando un mayor esfuerzo a pesar de la dominancia estocástica de primer orden. La caracterización (4) muestra que el esquema de incentivo óptimo siempre es monótono en la razón de verosimilitud  $l(x)$  y, por lo tanto, monótono en  $x$  si y sólo si la razón de verosimilitud es monótona en  $x$ . Supóngase que la densidad tiene dos jorobas y que la diferencia  $e_H - e_L$  es lo suficientemente pequeña para que las dos funciones de densidad se crucen más de una vez (las jorobas están entrelazadas). Esto da lugar a una razón de verosimilitud que no es monótona, lo que implica que existen dos valores  $x$  tales que el valor mayor tiene una razón de verosimilitud por debajo de uno, que habla a favor de bajo esfuerzo, mientras que el valor más bajo tiene una razón de verosimilitud superior a uno que sugiere un gran esfuerzo. En línea con la inferencia, al agente se le paga más por el valor más bajo que el valor más alto. Uno puede pensar en las dos jorobas como dos estados de la naturaleza: tiempos malos y buenos. El resultado más alto sugiere que fue obtenido el buen estado, pero condicionado a un buen estado, la evidencia sugiere que el agente holgazaneó.<sup>9</sup>

Se puede eludir este resultado empíricamente inverosímil suponiendo que el agente puede destruir el producto, en cuyo caso sólo son relevantes los incentivos no decrecientes (Innes 1990). O puede suponerse que la razón de verosimilitud es monótona, una propiedad común de muchas funciones de distribución y un supuesto utilizado frecuentemente tanto en estadística como en economía (Milgrom 1981). Pero la caracterización en (4) deja en claro que el modelo de esfuerzo básico no puede explicar formas comunes de incentivos. El uso universal de salarios a destajo, por ejemplo, no puede deberse a razones de verosimilitud similares.

El punto de vista inferencial también nos ayuda a entender un ejemplo problemático de Mirrlees ([1975] 1999). Mirrlees estudia la función de producción aditiva  $x = e + \varepsilon$  donde  $\varepsilon$  está distribuida normalmente con media cero y varianza constante. En otras palabras, el agente elige la media de una distribución normal con varianza constante. Analiza el problema con una elección continua de esfuerzo, pero la intuición principal puede transmitirse con sólo dos acciones.

Así, supóngase que el agente puede elegir entre un esfuerzo bajo (L) o alto (H) con un diferencial de costo (o de utilidad) igual a  $c > 0$ . Nótese que el cociente de verosimilitud  $f_L(x)/f_H(x)$  con la distribución normal tiende al infinito a medida que  $x$  va al infinito negativo. Por lo tanto, el lado derecho de (4), que se supone caracteriza la solución óptima, debe ser negativo para valores suficientemente pequeños de  $x$ , porque los valores de  $\lambda$  y  $\mu$  son estrictamente positivos cuando se implementa un esfuerzo alto como se discutió anteriormente. Pero eso es inconsistente con que el lado izquierdo sea estrictamente positivo

<sup>9</sup> Grossman y Hart (1983) muestran que lo único que se puede decir en un modelo general con resultados discretos  $x$  es que el esquema de incentivos óptimo no puede ser tal que siempre que al agente se le paga más, al principal se le pague menos. Debe haber un incremento en  $x$  tal que los pagos se muevan conjuntamente.

para todo  $x$ , independientemente de la función  $s_H(x)$ . ¿Cómo puede ser? Mirrlees conjeturó y demostró que esto debe ser porque la mejor solución puede ser aproximada dentro de un entorno arbitrario, pero jamás alcanzada.<sup>10</sup>

En particular, considérese un esquema  $s(x)$  que paga al agente una cantidad fija por encima de un punto de corte  $z$ , por ejemplo, y castiga al agente por una cantidad fija si  $x$  cae por debajo de  $z$ . Si la función de utilidad no está acotada inferiormente, Mirrlees muestra que se puede construir una secuencia de estándares cada vez más permisivos  $z$ , cada uno con un castigo cada vez más severo que mantenga el incentivo del agente a trabajar duro y la disposición del agente a aceptar todos los contratos en la secuencia. Además, esta secuencia puede elegirse de modo que la utilidad esperada del principal se aproxime al resultado de óptimo primario.

Una intuición de este resultado paradójico es que, aunque uno no pueda establecer  $\mu = 0$  en (4), uno puede acercarse a esta caracterización haciendo que  $\mu$  tienda a cero en una sucesión de esquemas que asegure que el lado derecho permanezca positivo.

La lógica inferencial sugiere una explicación intuitiva. *A pesar de la forma en que la distribución normal aparece a simple vista, los eventos de cola son extremadamente informativos cuando se trata de juzgar si el resultado observado es consistente con que el agente elija un nivel de esfuerzo alto en lugar de uno bajo.* La razón de verosimilitud  $f_L(x) / f_H(x)$  tiende al infinito en la cola inferior, lo que implica que un resultado muy alejado en la cola es una evidencia abrumadora en contra de la distribución de  $H$ . Estadísticamente, la distribución normal es afín al ejemplo de la caja móvil discutido en conexión con los resultados de óptimo primario. Los castigos en la cola son efectivos, porque se puede estar cada vez más seguro de que la fuente fue  $L$ .

### III. Hacia Modelos más Realistas

#### A. *¿Por qué esquemas de incentivos lineales?*

Uno podría dejar a un lado el ejemplo de Mirrlees como extremadamente poco realista (lo cual es) y proceder a estudiar contratos lineales porque son simples y ampliamente utilizados. Pero esto frustraría el propósito de estudiar incentivos más generales. Una función de producción aditiva con un término de error normal es un ejemplo muy natural a estudiar. La solución poco realista nos impulsa a buscar una razón más fundamental para la linealidad que la simplicidad y a revisar los supuestos que subyacen al modelo básico principal-agente. Esto nos lleva al análisis contenido en Holmström y Milgrom (1987).

El óptimo en el modelo básico tiende a ser complejo, porque existe un desequilibrio entre el espacio de esfuerzo unidimensional del agente y el espacio de control infinito dimensional disponible para el principal.<sup>11</sup> Por lo tanto, el principal puede ajustar los incentivos utilizando cualquier trozo de información minuciosa disponible sobre el desempeño del agente, lo que resulta en esquemas complejos.

El esquema de bonificación que se aproxima al óptimo primario en el ejemplo de Mirrlees funciona por la misma razón. Está sintonizado finamente porque se supone que el agente controla sólo el promedio de la distribución normal. Si enriquecemos el espacio de elección del agente permitiendo que el agente observe cierta información antes de elegir el esfuerzo, el esquema será de bajo rendimiento. Piénsese en un vendedor que recibe un bono discreto al final del mes si supera un objetivo de ventas. Supongan que puede ver su

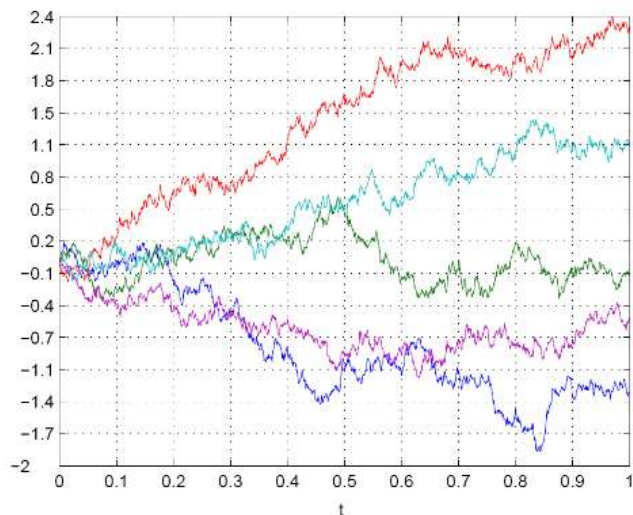
<sup>10</sup> Este resultado de inexistencia viola el supuesto de Grossman y Hart (1983) de que las probabilidades son estrictamente positivas.

<sup>11</sup> Para cada  $x$ , hay un control separado  $s(x)$ . El espacio de todas las funciones posibles  $s$  es de dimensión infinita.

progreso en el tiempo y ajustar su esfuerzo de ventas en consecuencia. *Luego, si el mes está a punto de terminar y todavía está lejos de alcanzar el objetivo de ventas, puede tratar de pasar las ventas potenciales al mes siguiente y obtener una ventaja para el bono del próximo mes. O, si ya ha alcanzado el objetivo, puede hacer lo mismo.* Este tipo de juego es común (ver Healy 1985 y Oyer 1998). En la medida en que tal juego sea disfuncional, se debe evitar un esquema de bonificación.

**Los esquemas lineales son robustos a la manipulación.** Independientemente de cuánto haya vendido el vendedor hasta la fecha, el incentivo para vender una unidad más es el mismo. Siguiendo esta lógica, armamos un modelo dinámico discreto donde el agente vende una o ninguna unidad en cada período y elige el esfuerzo en cada período en función del historial completo hasta ese momento. El principal paga al agente al final, basando el pago en la ruta completa de rendimiento. Suponiendo que el agente tenga una función de utilidad exponencial y el esfuerzo se manifieste como un costo de oportunidad (el costo es financiero), es óptimo pagar al agente de la misma manera en cada período: un salario base y una bonificación por venta.<sup>12</sup> La elección óptima de esfuerzo del agente también es la misma en cada período, independientemente de la historia pasada. Por lo tanto, el pago de incentivo óptimo al final del plazo es lineal en las ventas totales, independientemente de cuándo se produjeron las ventas.<sup>13</sup>

El modelo discreto tiene un análogo de tiempo continuo donde el esfuerzo instantáneo del agente controla la deriva de un **movimiento browniano con varianza fija** a lo largo del intervalo de tiempo  $[0,1]$ . El **modelo browniano** puede ser considerado como límite de un modelo de tiempo discreto bajo supuestos adecuados.<sup>14</sup> Es óptimo que el principal pague al agente una función lineal de la posición final del proceso (las ventas agregadas) en el momento 1 y que el agente elija una tasa de deriva constante (esfuerzo) en cada instante. Dado esto, la posición final del proceso en  $t = 1$  está normalmente distribuida, con la selección de esfuerzo constante del agente determinando la media y siendo constante la varianza. En otras palabras, volvemos al ejemplo de Mirrlees  $x = e + \varepsilon$  con  $\varepsilon$  normalmente distribuida y el agente eligiendo  $e$ . La única diferencia es que podemos limitar la elección del principal a un esquema de incentivo lineal con la pendiente óptima y el término constante a ser determinado.



Cinco simulaciones de un movimiento browniano (el eje  $x$  es el tiempo  $t$ , el eje  $y$  es  $Z(t)$ )

Fuente V. Granville *A Simple Introduction to Complex Stochastic Processes*, Diciembre de 2017

<sup>12</sup> La función de utilidad del agente es  $u(m, e) = 1 - \exp[-r(m - c(e))]$ , donde  $r$  es la aversión al riesgo absoluta del agente,  $c(e)$  es el costo de oportunidad del esfuerzo y  $m$  es dinero. Esta función de utilidad es multiplicativamente separable en lugar de serlo aditivamente. La razón por la que elegimos esta función es que los niveles de ingreso no afectan la elección del agente.

<sup>13</sup> Esto parece violar el principio de informatividad: el ritmo de ventas no es utilizado. La razón es que resulta óptimo para el principal implementar una acción constante y con esta implementación el ritmo de ventas es irrelevante.

<sup>14</sup> Ver Holmström y Milgrom (1987) para una derivación intuitiva y Hellwig y Schmidt (2002) para un análisis riguroso.

Notablemente, al enriquecer juiciosamente la elección del agente, hemos llegado a una solución muy simple del problema de Mirrlees.<sup>15</sup>

### B. Extensiones y Enfoques Alternativos

Hay una literatura emergente que estudia la solidez de los esquemas de incentivos lineales, donde la solidez es definida como el resultado máximo-mínimo en un entorno más amplio, que puede o no ser conocido por el principal. Hurwicz y Shapiro (1978) es el primer artículo en esta línea que muestra que la división lineal 50-50 ampliamente observada en la aparcería se puede racionalizar dentro de esta línea. Diamond (1998), Edmans y Gabaix (2011), Chassang (2013) y Carroll (2015) son variantes más recientes sobre este tema. El artículo de Carroll es especialmente elegante y, en el fondo, lo suficientemente simple como para aplicarse en entornos económicos más ricos. El documento capta la robustez en un sentido fuerte (una rentabilidad mínima garantizada para el principal en un entorno ampliamente desconocido), pero no parece tan manejable como el modelo de Holmström-Milgrom, al menos no todavía.

Yuliy Sannikov ha avanzado mucho y popularizado el uso de modelos de tiempo continuo para estudiar problemas de incentivos que son tratables así como relevantes. Sannikov (2008) resuelve un problema general de agencia no estacionario usando técnicas poderosas de la teoría del control estocástica. Estas técnicas requieren buen ojo para encontrar los supuestos correctos que permitan hacer avanzar el análisis, pero la recompensa puede ser alta como lo ilustra el modelo dinámico de compensación de CEOs de Edmans et al. (2012). El modelo presenta un caso convincente para usar cuentas dinámicas de incentivos que mantienen los fondos en custodia y ajustan la proporción de deuda y patrimonio en respuesta a la información entrante. Intuitivamente, y también en la práctica, la solución tiene mucho sentido. Otro ejemplo de un modelo de tiempo continuo de interés en la realidad es DeMarzo y Sannikov (2006).



[Yuliy Sannikov](#) (Stanford Business)  
Sebastián Di Tella & Yuliy Sannikov,  
[Optimal Asset Management Contracts with  
Hidden Savings](#) (Nov., 2016)

### C. Modelo Lineal con Una Tarea

Es sencillo resolver el ejemplo de Mirrlees con un esquema de incentivo lineal  $s(x) = ax + \beta$ . El pago del agente se distribuye normalmente con media  $ae + \beta$  y varianza  $\alpha^2\sigma^2$ . Dada una función de utilidad exponencial, la utilidad del agente se puede escribir en términos de su equivalente cierto:

$$CE_A = \alpha e + \beta - \frac{1}{2} r \alpha^2 \sigma^2.$$

El agente tiene preferencias de media-varianza donde  $r$  es el costo de soportar riesgos. El agente elige  $e$  para que el costo marginal del esfuerzo sea igual al retorno marginal:  $c'(e) = \alpha$ .

<sup>15</sup> Cuando el agente puede observar su progreso y hacer su elección de esfuerzo contingente al estado actual, el agente puede generar esencialmente cualquier distribución de la posición final del proceso en  $t = 1$  usando un puente browniano (que toma al proceso desde su punto de inicio en  $t = 0$ , a un punto arbitrario en  $t = 1$ ). Agradezco a Michael Harrison haber mostrado esto.

El principal es neutro al riesgo por lo que su equivalente cierto es:

$$CE_P = (1 - \alpha) e - \beta.$$

Como el modelo es uno de utilidad transferible, primero se puede resolver la pendiente óptima maximizando la suma de los dos equivalentes  $CE_A + CE_P$ . Esto proporciona el máximo excedente global que luego se divide entre las partes usando  $\beta$ . Usando la condición de primer orden para la elección  $e$  del agente, obtenemos entonces el valor óptimo de

$$(6) \quad \alpha^* = [1 + r \sigma^2 c'']^{-1},$$

donde la dependencia de  $c''$  de  $e$  ha sido omitida ( $c''$  es constante si la función de costo es cuadrática).

La lógica de este modelo es refrescantemente simple. El agente trabaja más duro cuanto más fuerte sea el incentivo (mayor  $\alpha$ ). De acuerdo con (6) la fuerza de incentivo óptima  $\alpha^*$  siempre cae entre cero (sin incentivo) y uno (incentivo óptimo primario), debido al riesgo que el agente tiene que soportar. Si el agente es más reacio al riesgo o la medición del rendimiento es menos precisa, el incentivo financiero es más débil. ¿Qué hay de  $c''$ ? De la condición de primer orden del agente obtenemos  $1/c'' = de/da$ . La derivada mide la capacidad de respuesta del agente a un mayor incentivo. El agente es más receptivo a los incentivos si la función de costo es menos empinada (menor  $c''$ ), lo que resulta en una tasa de comisión más alta como se aprecia en (6).

La acción unidimensional del agente y el control unidimensional del principal están apareados, lo que hace que este modelo se comporte muy bien. Se puede extender el modelo en muchas dimensiones. Se pueden estudiar los costos y beneficios de diferentes proyectos elegidos conjuntamente, tecnologías de producción y sistemas de monitoreo, por ejemplo, y obtener respuestas simples. La variante más interesante en tales ejercicios mentales se refiere a la función de costo de oportunidad  $c(e)$ . Hay muchas maneras en que el principal puede cambiar la función de costo de oportunidad del agente para reducir el costo de los incentivos. Esta idea fue central en iniciar mi trabajo sobre multitareas con Paul Milgrom (Holmström y Milgrom 1991, 1994).

#### IV. Multitareas

La multitarea, la realidad de que el trabajo de un agente consista en muchas tareas, generó un cambio importante en la mentalidad y el enfoque con los que se encara el problema principal. En lugar de estudiar cómo conseguir que el agente trabaje lo suficiente en una sola tarea, la atención se concentró en cómo el agente asigna su esfuerzo a las tareas de manera alineada con los objetivos del principal. Cuando las tareas son interdependientes, el diseño óptimo debe considerar los incentivos del agente en su totalidad. Conocer la cartera completa de actividades del agente - cuáles son su autoridad y sus responsabilidades - resulta esencial para diseñar una solución coherente y equilibrada que tenga en cuenta las interdependencias. Esto es desafiante cuando las actividades fáciles y difíciles de medir compiten por la atención del agente o si las medidas de desempeño disponibles están mal alineadas con los objetivos del principal.

##### A. Tareas fáciles versus difíciles de medir

Consideren el caso de un agente con dos tareas. Una tarea puede medirse perfectamente: piensen en la cantidad vendida. La otra tarea es muy difícil de medir: piensen en la reputación de la empresa. Para esta última puede haber algunas medidas disponibles, por ejemplo, comentarios de los consumidores. Pero tal información es selectiva y a menudo parcial. Las personas con experiencias de ventas infelices probablemente se quejen más a

menudo que las personas con experiencias felices. Algunos clientes pueden haber tenido un mal día. Y algunos clientes importantes pueden no tener tiempo para hacer comentarios. Todo esto hace que sea difícil evaluar cómo los comentarios de los consumidores se traducen genuinamente en una reputación valiosa en el futuro.

Para ser un poco más formal acerca de esto, podemos usar una extensión multitarea del modelo lineal (ver el Apéndice para una versión general). Hacemos que el rendimiento en cada tarea se mida por separado, la cantidad mediante  $x_1 = e_1$  y la reputación mediante  $x_2 = e_2 + \varepsilon_2$ . No hay un término de ruido  $\varepsilon_1$  porque supongo que la cantidad se puede medir perfectamente. La varianza de  $\varepsilon_2$  es mayor cuanto más ruidosa sea la respuesta del consumidor.

Sea el objetivo del principal  $B(e_1, e_2) = p_1 e_1 + p_2 e_2$ , donde  $p_1$  y  $p_2$  miden el valor de cantidad y reputación, y supóngase por un momento que la función de costo del agente sea separable:  $C(e_1, e_2) = c_1(e_1) + c_2(e_2)$ . En este caso, los incentivos del agente también se pueden analizar por separado y es óptimo establecer cada coeficiente de forma independiente de acuerdo a la fórmula para el caso de una sola tarea. La tasa de comisión en  $x_1$  será  $\alpha_1 = p_1$ , proporcionando incentivos de óptimo primario para la elección de cantidad, mientras que la tasa de comisión en  $x_2$  será  $\alpha_2 < p_2$ , con un trade-off de riesgo versus incentivos para un resultado de óptimo secundario.

Ahora, supóngase que las tareas son sustitutos: cuanto más tiempo el agente invierte en una tarea, mayor es su costo marginal de invertir tiempo en la otra tarea (la derivada parcial cruzada  $c_{12} > 0$ ). En este caso, ya no es óptimo proporcionar al agente incentivos de óptimo primario para elegir cantidad ( $\alpha_1 = p_1$ ), porque bajar  $\alpha_1$  ligeramente en el margen no costará nada en el valor perdido en la primera tarea, pero será estrictamente beneficioso proporcionando incentivos en la segunda tarea. Al agente se le puede dar el incentivo de suministrar la misma cuantía de calidad con un  $\alpha_2$  más bajo, lo que reducirá el costo del riesgo. O, alternativamente, cuando  $\alpha_1$  es menor, el costo marginal de invertir tiempo en calidad habrá disminuido, lo que hará que el agente dedique más tiempo a esa tarea. De cualquier forma, disminuir el incentivo en cantidad es ventajoso para el suministro de calidad; por cuánto, dependerá de la precisión con que se pueda medir la calidad y cuán sustituibles sean las tareas en la función de costos. Si las dos tareas son sustitutos perfectos (es decir, el agente sólo asigna su tiempo entre las dos tareas, entonces la función de costo es  $C(e_1 + e_2)$ ) y la atención a la calidad es esencial (esto requerirá una función de beneficio no lineal  $B$ ), cualquier incentivo a la cantidad tendrá que coincidir con un fuerte incentivo correspondiente a la calidad, de modo que  $\alpha_1 = \alpha_2$ . Esto hará que el agente esté indiferente entre invertir tiempo en cualquier tarea y elegirá la asignación que resulte mejor para el principal. El costo es que el incentivo para ambas tareas deberá reducirse porque la calidad no es adecuadamente medida. Si no hay ninguna medida de calidad (es decir, la varianza de  $\varepsilon_2$  es infinita), entonces ningún incentivo para ninguna de las tareas será óptimo. Si  $c'(0) < 0$ , el agente aún elegirá un nivel positivo de esfuerzo total.

### B. Desalineación y Manipulación

La desalineación es una variación importante en el tema multitarea (Baker 1992, 2002). También puede ser analizada con el modelo general de multitareas.

Supongamos nuevamente que el valor del principal es  $B(e_1, e_2) = p_1 e_1 + p_2 e_2$ , y que sólo hay una medida de rendimiento  $x(e_1, e_2) = g_1 e_1 + g_2 e_2$ . Aunque no hay incertidumbre en la medida, existe un problema de incentivo no trivial si los vectores  $p$  y  $g$  no están alineados. Por ejemplo, supongamos  $x = e_1 + e_2$  pero  $B(e) = e_1$ . El principal sólo valora el esfuerzo en la tarea principal 1, pero el agente puede producir el producto medido  $x$  con  $e_1$  y  $e_2$ . La segunda actividad puede interpretarse como "manipulación", que el agente puede sentir como una aversión moral, pero si el principal presiona demasiado sobre el rendimiento medido  $x$  con la esperanza de lograr que el agente trabaje duro en la tarea principal, el

agente se ocupará también de manipular. El principal termina compensando al agente por un rendimiento inútil, por lo que la desalineación causa despilfarro, más aún cuanto mayor sea  $\alpha$ . Será óptimo establecer  $\alpha < 1$ , porque con  $\alpha = 1$  el costo marginal de reducir  $\alpha$  es estrictamente negativo ( $e_1$  es óptimo primario, mientras que  $e_2$  no).

Ambas variantes de multitarea, las disparidades en los errores de medición y la desalineación entre las medidas de rendimiento y el valor creado son variantes relevantes del modelo general multitarea. Dependerá del contexto cuál sea la más natural de usar.

### C. “Lo Barato Sale Caro”

Una manipulación general de medidas de desempeño estuvo detrás del reciente **escándalo de Wells Fargo** (Tayan 2016). Wells Fargo había evitado los escándalos bancarios asociados con la crisis financiera. Era conocido por su prudencia en los préstamos, obteniendo ganancias poniendo énfasis en su banca minorista y el servicio al cliente. Sus gerentes de sucursales estaban muy incentivados por ventas cruzadas: a saber, lograr que sus clientes bancarios habituales compraran una gama de productos, como líneas de crédito. Esta parte de su negocio bancario había crecido constantemente y había sido altamente rentable. Pero el crecimiento continuo también requería nuevos clientes y, eventualmente, a medida que los objetivos de ventas se iban estrechando (el rendimiento de los gerentes de sucursales era diariamente medido), algunos de los gerentes abrieron nuevas cuentas para sus clientes sin el conocimiento de los mismos. Las **cuentas de shell** eran como una segunda actividad en el modelo de dos tareas descrito anteriormente. Todo esto mejoró el rendimiento medido y generó bonificaciones, pero como no había actividad real en las cuentas, esta actividad generó ganancias minúsculas para Wells Fargo (U\$S 2,6 millones según Tayan 2016). Las cuentas de Shell les causaron costos mínimos a los clientes (estimativamente, U\$S 2.50 por cuenta), pero por supuesto eran ilegales. Finalmente, la estafa fue descubierta, lo que provocó el despido o renuncia de miles de empleados y, finalmente, también la renuncia del CEO Mr. Stumpf. Wells Fargo tuvo que pagar U\$S 185 millones en multas, pero el mayor costo fue el enorme daño a su reputación estelar.<sup>16</sup>

Las explicaciones del **derrame de petróleo de BP en el Golfo de México** apuntan en muchas direcciones, pero el hecho de que BPX -el brazo de exploración de BP- fuera alentada a ser más agresiva en su actividad de exploración fue probablemente uno de los culpables (Garicano y Rayo 2016). Fueron impulsados los resultados medibles (implícita o explícitamente), comprometiendo la seguridad. Este es un ejemplo donde hay muchas actividades, algunas de fácil medición (exploración exitosa) y otras que no lo son. Si bien la seguridad puede ser monitoreada, no resulta sencillo realizar un monitoreo intenso de si las reglas se están siguiendo. Leves demoras en el servicio o en la reparación de piezas defectuosas, especialmente en el caso de sistemas de respaldo y chequeo, pueden parecer entrañando un riesgo mínimo y, por lo tanto, estar sujetas a compromisos en situaciones bajo presión.

Hay varios escándalos asociados a la introducción de **incentivos a los maestros basados en pruebas**. Uno de los más notorios ocurrió en el sistema de escuelas públicas de Atlanta, donde los resultados en las Pruebas de Competencia mostraron irregularidades en los puntajes de las pruebas (extraordinarios aumentos y disminuciones en un solo año). Una investigación reveló que 44 de 56 escuelas engañaron en el Examen de Competencia de 2009, y 178 maestros y directores fueron hallados culpables de corregir las respuestas de los estudiantes. En 2015, 11 maestros fueron condenados por fraude organizado y recibieron sentencias severas, incluyendo penas de prisión.

<sup>16</sup> Por lo que la segunda actividad  $e_2$  en realidad le costó mucho al principal y, por lo tanto, debe ingresar como  $B(e)$ , pero con un gran signo negativo, reflejando los costos de pérdida de reputación.

El escándalo fue consecuencia involuntaria de la Ley *No Child Left Behind* de 2001 e ilustra cómo las buenas intenciones pueden salir mal. Como en el caso de Wells Fargo y en otros similares, los culpables de actos delictivos describen sus actos como resultado de tener una presión excesiva por el desempeño.

El debate sobre los costos y beneficios de evaluar y premiar a docentes y escuelas está en curso (Neal 2011). El problema no es tanto acerca de hacer trampa, sino que se enfoca más acerca de otros costos, como *enseñar la prueba* [*teaching to the test*]. Al igual que Stephen Kerr (1975), en su artículo ya mencionado *On the Folly of Rewarding A, While Hoping for B*, el modelo multitarea advierte contra el uso agresivo de premios para temas fáciles de medir como las matemáticas a expensas del desarrollo de habilidades sociales importantes y otras difíciles de medir. Pero la economía va más allá de la crítica de Kerr de recompensar metas estrechas al estudiar muchos instrumentos sustitutivos que las empresas pueden usar con bastante eficacia. Volveré al punto luego de resumir las lecciones principales del problema de tareas múltiples.<sup>17</sup>

#### D. *Incentivos de Baja Potencia*

Una primera lección de la multitarea es que hay dos formas de proporcionar incentivos para una actividad difícil de medir, como la oferta de calidad. El enfoque *directo* es pagar por la calidad medida, por mala que sea esa medida. El *indirecto* es reducir el incentivo sobre la cantidad, porque esto reduce el costo de oportunidad de proporcionar calidad. La eficacia de la forma indirecta depende de qué tan pobre sea la medida de calidad y de cuán fuertemente las dos tareas compitan entre sí por el tiempo del agente. En el caso puro de asignación de tiempo, las tareas compiten directamente entre sí por la atención del agente, dando lugar a un efecto máximo de sustitución.

Una segunda lección es que los problemas de medición pueden explicar por qué *incluso las actividades perfectamente medidas pueden quedar sin incentivos*. El hecho de que el agente pueda controlar bien una medida, porque la medida es precisa con respecto a una tarea o un subconjunto de tareas, no implica que deba usarse. A veces, sucede todo lo contrario. A los empleados les suelen gustar las medidas contables que pueden controlar, como los costos, pero estas medidas a menudo son parciales y, por lo tanto, están mal alineadas con la verdadera creación de valor, en cuyo caso se requieren incentivos de baja potencia para estas medidas.

La tercera lección es la más general. Como el costo de oportunidad de proporcionar incentivos para cualquier tarea dada depende de los incentivos para todas las otras tareas, siempre se debe *considerar la cartera completa de tareas de los agentes al diseñar los incentivos*.

#### V. Instrumentos de Incentivo Alternativos de las Empresas

Las empresas usan esquemas de pago por desempeño con moderación. La mayoría de los empleados tienen un salario fijo. Williamson (1985) ofrece una explicación: la integridad de las medidas de desempeño dentro de la empresa es más débil que en el mercado. La multitarea ofrece una explicación complementaria. Las empresas tienen acceso a muchos sustitutos de los incentivos de pago por rendimiento que no son fácilmente accesibles por medio del mercado. El más importante es la capacidad de controlar el trabajo mediante

<sup>17</sup> También hay ejemplos en los que las empresas implementan planes de valores a destajo agresivos con éxito. Safelite, la empresa dominante en el mercado de reemplazo de parabrisas, es un caso puntual (Lazear 2000). Lincoln Electric, un fabricante de equipos de soldadura, es otro (Milgrom y Roberts 1995). En esos casos, la introducción de salarios a destajo requirió una serie de cambios coincidentes en la estructura organizacional para evitar el tipo de problemas antes descritos.



asignaciones de trabajo, diseños de trabajo y una variedad de reglas implícitas y explícitas establecidas por la empresa.

### A. Reglas y Restricciones

Considérese el caso de trabajar en casa en lugar de ir a la oficina.<sup>18</sup> Supóngase que la tarea asignada al agente -una sola, por simplicidad- se puede realizar tan bien en casa como en la oficina. La única diferencia entre las dos opciones es que el agente tiene otras actividades que puede realizar en el hogar, actividades que brindan beneficios privados. Puede trabajar en otros proyectos en el hogar, hacer deportes, mirar televisión, descansar, jugar con sus hijos y demás. Podemos pensar en estas actividades de valor privado como compensaciones vis-a-vis el costo de trabajar para el principal. El inconveniente es que reducen el tiempo que el agente puede dedicar a la tarea del principal.

Si al agente se le pagara el valor de la producción que produce, no habría problemas de trabajar desde casa. Como demandante residual, el agente haría el cálculo socialmente correcto de cuánto tiempo dedicar a actividades privadas y cuánto a la tarea del principal. Pero este contrato también obligaría al agente a asumir riesgos. Si el riesgo es costoso y, por lo tanto, se reduce el poder de incentivo, la comisión del agente no reflejará por completo el valor marginal de sus actividades. La tarea del principal, como resultado, se volverá menos competitiva frente a las actividades privadas. Suponiendo que cada una de las actividades privadas pueda cancelarse o que el agente tenga la libertad de realizar la actividad como lo desee - una decisión de cero-uno sobre cada actividad - ¿qué actividades deberían permitirse y cuáles se cancelarían? Tengan en cuenta que permitir actividades privadas es una forma de pago al el agente. Por lo tanto, el principal también se beneficiará al permitir que el agente emprenda una actividad si el excedente total se incrementa de esa manera. Con una comisión fija, si se cancela una tarea privada, el agente invertirá el tiempo liberado en la tarea del principal. Entonces, si la tarea debe clausurarse es sólo una cuestión de si la cantidad elegida por el agente de una actividad privada, dada la tasa de comisión establecida por el principal, vale más cuando se gasta en la tarea del director.

A menor comisión recibida, menos tiempo invertirá el agente en la tarea del principal y más en sus tareas privadas. En consecuencia, el valor marginal de la tarea del principal aumenta mientras que los valores marginales en las tareas privadas del agente disminuyen a medida que disminuye la tasa de comisión: el agente pasa demasiado tiempo en tareas privadas y muy poco en las del principal. Entonces, excluir las actividades privadas se vuelve más importante, a menor tasa de comisión. Las restricciones en las actividades privadas actúan como sustitutos de los incentivos monetarios al aumentar la cantidad de tiempo que el agente gasta en la tarea del principal. En este ejemplo, las restricciones se colocan en orden decreciente de importancia de las tareas (medido por el valor privado medio de las tareas).

El incentivo del agente será más débil si el rendimiento en la tarea principal es medido con menor precisión (e.d., la varianza de  $\varepsilon$  es mayor) y esto implicará que se excluyan más ta-

---

<sup>18</sup> Puede hallarse un tratamiento formal de este ejemplo en Holmström y Milgrom (1991).

reas para sustituir los incentivos más débiles.<sup>19</sup> Responsabilidad y libertad se mueven conjuntamente a medida que cambia la precisión de la medición.<sup>20</sup>

Más allá de las restricciones de tareas, otros instrumentos también pueden ser importantes. Por ejemplo, las reglas burocráticas en las organizaciones son tan frecuentes como las quejas sobre ellas. La burocracia es detestada porque desalienta la iniciativa y la innovación. Muy a menudo la gente no percibe que hay un propósito para la mayoría de las reglas burocráticas. En primera instancia, estas reglas no son síntomas de una enfermedad de la organización, sino que forman parte de la solución a problemas complicados de incentivos. Debido a que el pago por rendimiento tiende a ser un instrumento de incentivo costoso en la empresa, debido a problemas con la medición del desempeño, las reglas y restricciones burocráticas a menudo servirán como sustitutos efectivos (Holmström 1989).

Las empresas también fijan diversas otras restricciones sobre sus empleados que limitan lo que pueden hacer. Una de las reglas más obvias y universales (hasta hace poco, tal vez) es el requisito de que un empleado no puede trabajar para otras empresas. Otro requisito común es que los empleados deben venir a la oficina por un horario determinado. Naturalmente, estas restricciones también son impulsadas por otras consideraciones, pero su efecto sobre los incentivos de los empleados no es trivial.

### *B. Diseño del Trabajo*

Otra dimensión importante del diseño del trabajo es la distribución de tareas entre los trabajadores.

Regresemos al caso de dos tareas con un agente. Una tarea es medida perfectamente (una tarea de rutina) y la otra lo es con un ruido considerable (una tarea innovadora). La función de costo del agente depende únicamente del tiempo total dedicado. Concluí que proporcionar incentivos de baja potencia en la tarea de rutina es apropiado como un incentivo para la innovación.

¿Qué pasaría si se pueden dividir las dos tareas entre dos agentes? Un agente se ocuparía de la tarea de rutina mientras que el otro se especializaría en la innovadora. Esto evita la tensión entre tareas rutinarias e innovadoras. La tarea de rutina se puede proporcionar con incentivos de gran potencia, mientras que la tarea innovadora tiene incentivos de baja potencia. De manera más general, se puede demostrar (Holmström y Milgrom 1991) que si hay un continuo de tareas (para evitar problemas con números enteros) con costos de medición variables, entonces se deberían asignar todas las tareas difíciles de medir (definidas mediante un punto de corte) a un agente y las tareas fáciles de medir al otro agente. Las

<sup>19</sup> Los resultados de estática comparativa como éste, que involucran diversas variables endógenas, requerirán, en general, un análisis más complejo utilizando métodos monótonos (ver Milgrom y Roberts 1990 y Holmström y Milgrom 1994). Como la exclusión de tareas privadas sólo depende de la tasa de comisión, pero no directamente del error de medición, este caso es sencillo. Si, en cambio, el valor de la tarea del principal aumenta, existe un efecto directo negativo sobre la exclusión, y por lo tanto es posible que se permitan menos tareas privadas aunque aumente la tasa de comisión. Hay dos maneras de aumentar la atención del agente hacia la tarea del principal: aumentar la tasa de comisión o excluir tareas privadas. Ambos instrumentos pueden usarse conjuntamente.

<sup>20</sup> En un artículo importante, Prendergast (2002) observa que, empíricamente, a menudo vemos que el riesgo más alto se combina con incentivos más fuertes, no más débiles, como se acaba de describir. La razón para cambiar la conclusión es que el riesgo en su modelo se refiere a la productividad, no al error de medición, y por lo tanto, es valioso que el agente responda a las circunstancias cambiantes. Este caso se puede incorporar en el modelo multitarea reemplazando la función de producción aditiva por una multiplicativa. La acción del agente será entonces una estrategia contingente y la medida única será un promedio ponderado de los posibles estados de la naturaleza. El modelo de Prendergast y el modelo multitarea pueden explicar el co-movimiento positivo (ver Baker y Jorgensen 2003).

tareas difíciles de medir se proporcionan con incentivos iguales de baja potencia, mientras que las tareas fáciles de medir reciben incentivos de alta potencia iguales. El agente más intolerante al riesgo asume las tareas de bajo riesgo.

Más especialización puede ser una forma de lidiar con el dilema de la enseñanza de la prueba que se debate tan acaloradamente [V. *pág. 16*]. La idea ha sido promovida por Neal (2011) en base al modelo multitarea. Por supuesto, ya existe un grado significativo de especialización en la enseñanza, pero es posible que se desee aumentar y trasladar la responsabilidad de las habilidades interpersonales [*soft skills*] a los profesores que se especializan en éstas. Estos docentes no deberían ser considerados responsables por resultados de exámenes estrechos.

Hay una importante **lección adicional oculta en la lógica de la especialización**. En el modelo discutido, no se comparte ninguna tarea porque eso requeriría incurrir en riesgo por ambos agentes en la tarea dada. Si un agente ya está incentivado para hacer la mitad del proyecto, incurriendo en el costo de riesgo asociado con los incentivos adecuados, no hay costo de riesgo adicional para que él haga toda la tarea. Si bien este resultado depende en parte de nuestra función de producción aditiva y del continuo de tareas, la lógica es clara. Hay un costo fijo para que un segundo agente se una al proyecto, porque ambos tienen que ser incentivados. Esta característica tiene una validez más general y puede ser un fundamento de la **responsabilidad individual (unidad de mando)**.

El mismo principio es relevante para la **rotación de trabajos**. Un beneficio de la rotación laboral es que ofrece una forma de evaluación relativa del desempeño: tener dos agentes trabajando en la misma tarea proporciona información sobre la dificultad de la tarea y esto es valioso para evaluar el desempeño de cada agente de acuerdo con el principio de informatividad. Pero también hay un inconveniente desde el punto de vista del incentivo. Si la información de rendimiento entra con un retraso, uno quiere hacer que el incentivo del primer agente dependa de los resultados que surjan durante el turno del segundo agente. Esto es similar a dos agentes que comparten la misma tarea que se discutió anteriormente. El primer agente está parcialmente incentivado para los períodos posteriores debido a la información retrasada, y desde este punto de vista sería eficiente dejarlo continuar.

### *C. Incentivos Profesionales*

La gente busca ser apreciada por su trabajo. Los modelos de interés profesional (Holmström 1999a; Dewatripont, Jewitt y Tirole 1999) pueden capturar el deseo de apreciación a través de un mecanismo dinámico de señalización. Los empleadores van aprendiendo sobre el valor de los empleados a partir del desempeño pasado. El valor inicialmente se desconoce (también para el empleado) pero se revelará al menos parcialmente con el tiempo. Los empleados tratan de causar una impresión favorable porque un mejor rendimiento puede generar promociones, un salario más alto, más estatus y otros beneficios en el futuro.

Se puede esperar que las preocupaciones laborales sean más poderosas en las empresas que en los mercados, porque en las empresas los empleados generalmente saben a quién impresionar. Las empresas pueden explotar esta motivación al indicar lo que se desea de un empleado. En las encuestas a los empleados, una de las quejas más comunes es que los empleados desearían una comprensión más clara de lo que se espera de ellos. Si supieran lo que quiere el jefe, podrían trabajar para causar una buena impresión en esas dimensiones. El ansia de apreciación y el deseo de impresionar a los superiores explica por qué un simple cambio en el sistema de contabilidad puede tener un gran impacto en el comportamiento de los empleados. Los indicadores clave de rendimiento son importantes porque indican lo que la gerencia quiere. Solos o combinados con bonos relativamente modestos, pueden tener un efecto sorprendentemente fuerte. Poner dinero detrás de una medida

transmite un mensaje más fuerte de lo que se espera. Los empleados también responden a enunciados de misión como en Dewatripont, Jewitt y Tirole (1999).

Los incentivos de carrera pueden ser tan fuertes como para obviar cualquier necesidad de incentivos financieros. De hecho, pueden ser demasiado fuertes y llevar a actividades desmedidas de alcahuetería e influencia (Milgrom y Roberts 1988, Holmström y Ricart i Costa 1986; Prendergast 1993). Al igual que los incentivos financieros, los incentivos profesionales pueden desequilibrarse, ya que algunos indicadores de rendimiento son más visibles y, por lo tanto, más destacados que otros. La empresa puede regular la visibilidad a través del acceso que el empleado tiene a los superiores y puede instituir prácticas de pago y promoción que no respondan al rendimiento con tanta fuerza. Basar los aumentos salariales y las promociones en la antigüedad silencia los incentivos profesionales que se perciben como demasiado fuertes o que conducen a una conducta miope. En general, la jerarquía general de la empresa y sus políticas de promoción son poderosos impulsores de incentivos sin contrapartidas directas en el mercado.<sup>21</sup>

Los problemas de la multitarea tienden a manifestarse de la misma manera cuando las preocupaciones de carrera son el conductor del comportamiento de los empleados en lugar de las bonificaciones explícitas. Esta es la razón por la cual la discusión anterior sobre la multitarea, asumiendo que las comisiones son instrumentos de incentivo, proporciona una guía útil a pesar de que las comisiones son relativamente menos utilizadas por las empresas.

#### VI. Dos Sistemas de Incentivo: Empleo versus Contratos

La discusión anterior aclara el punto de que las empresas tienen acceso a una variedad de instrumentos de incentivos que pueden compensar los incentivos de bajo rendimiento de pago por desempeño. En esta sección, quiero resaltar dos implicancias para el diseño y uso de sistemas de incentivos.

El primero es que la estructuración de incentivos eficientes dentro de la empresa requiere un uso concertado de todos los instrumentos de incentivo disponibles. Esto es paralelo a la lección anterior de la multitarea de que las comisiones para todas las tareas deben considerarse en forma concertada. El segundo punto es que los incentivos en empresas y en mercados forman dos sistemas lógicamente coherentes, cada uno con su clara ventaja comparativa.

En Holmström y Milgrom (1994) ilustramos ambos puntos al aplicar nuestro modelo multitarea a la elección entre empleo y contratación. Queremos explicar los hallazgos de Anderson y Schmittlein (1984) y Anderson (1985), quienes estudiaron cómo se organiza la venta industrial en la industria electrónica. Algunos agentes de ventas trabajan como representantes independientes, otros como empleados. Una empresa a menudo usa ambas formas de organización de ventas. ¿Qué determina la elección entre las dos formas de ventas y cómo se estructuran los incentivos dentro de cada alternativa?

Anderson y Schmittlein encuentran que las medidas más importantes que determinan la elección entre empleos y contratos independientes fueron "la dificultad de evaluar el rendimiento" y "la importancia de las actividades de no venta". La evaluación del desempeño es difícil para las ventas complejas, mientras que las actividades de no venta son importantes cuando se desea la cooperación con otros agentes de ventas. Los aumentos en ambas medidas hicieron más probable el empleo. Además, los representantes independientes eran compensados totalmente mediante comisiones y se les permitía representar a otros

<sup>21</sup> La literatura relacionada sobre contratación relacional estudia el alcance de los contratos informales que pueden sustentar reglas y prácticas; ver p. Levin (2003); MacLeod y Malcomson (1988); y Baker, Gibbons y Murphy (1994, 1999).

fabricantes, mientras que los empleados de ventas tenían salarios fijos y no podían vender los productos de otros fabricantes.

Estos hallazgos se pueden explicar en nuestro entorno multitarea estándar donde las tareas compiten por la atención del agente de ventas. El hecho de que los agentes independientes sean libres de vender los productos de otros fabricantes y tengan fuertes incentivos es paralelo al argumento anterior sobre darle a un agente la libertad de realizar tareas privadas si la tasa de comisión es alta, pero no si es baja. La dificultad de evaluar el rendimiento disminuye la tasa de comisión, al igual que el deseo de que el agente asigne tiempo a las actividades no de venta (cooperativas).

El modelo incluye dos tipos de actividades de venta: venta directa, que afecta los rendimientos a corto plazo y está influenciada por las comisiones, y venta indirecta que aumenta las ventas futuras. Suponemos que los últimos retornos solo se pueden transferir a través de la propiedad. Ésta es una forma reducida de introducir un rol para la propiedad (Grossman y Hart 1986). Un agente de ventas independiente posee los retornos a largo plazo; un empleado de ventas no. El remedio para la pérdida de incentivos a largo plazo es pagar al empleado de ventas un salario fijo. Esto hace que el empleado de ventas esté dispuesto a asignar atención a las tareas de la forma que desee la empresa, a costa de iniciativas generales más débiles.

Tenemos, entonces, dos sistemas coherentes de incentivos encontrados por Anderson y Schmittlein. Un contratista independiente (representante de ventas) posee los rendimientos a largo plazo, se le pagan altas comisiones por ventas directas y es libre de representar a otras empresas. Un empleado de ventas no posee los ingresos a largo plazo, recibe un salario fijo y no puede representar a otras empresas. Y de acuerdo con la evidencia, el empleo se ve favorecido si la venta directa es difícil de medir o si las actividades que no son de venta son valiosas.

Otra forma de introducir una distinción entre empleo y contratación es suponer que una empresa puede restringir las tareas de un empleado de ventas más fácilmente que a un representante independiente (Holmström 1999b). Ambas alternativas muestran que la *teoría de los contratos de incentivo y la teoría de los derechos de propiedad son complementarias*. Juntas pueden proporcionar una perspectiva más rica de la organización.

## VII. Conclusión

Permítaseme cerrar con un resumen del viaje intelectual que he descrito. Cuando comencé a estudiar riesgo moral, el paradigma principal era el modelo básico de esfuerzo unidimensional. A pesar de la aparente simplicidad de este modelo, se comportaba de maneras que a veces eran desconcertantes. El Principio de Informatividad reveló la lógica básica del modelo, proporcionando información útil sobre el valor de la información al mismo tiempo que dejaba en claro que el modelo básico no puede explicar la forma de los esquemas de incentivos comunes.<sup>22</sup> Esto nos llevó a Paul Milgrom y a mí a preguntar por qué los esquemas de incentivos son lineales. Nuestra disposición a escuchar el modelo básico de agencia nos introdujo, por casualidad, en el mundo de la multitarea, que abrió un amplio conjunto de problemas y oportunidades. El valor de los incentivos de poca potencia en el contexto de la multitarea explica por qué las empresas hacen tan poco uso de bonificaciones explícitas y en su lugar usan alternativas como el diseño de trabajo y reglas burocráti-

---

<sup>22</sup> Se prestó mucha atención a los problemas con el enfoque de primer orden: el hecho de que, en general, no se puede reemplazar la restricción de compatibilidad de incentivos del agente con una condición de primer orden. Rogerson (1985) proporcionó una condición suficiente para el enfoque de primer orden y los esfuerzos continúan para refinar su trabajo. Pero ahora es evidente que el modelo de esfuerzo unidimensional como tal tiene serias deficiencias.

cas para construir sistemas de incentivos coherentes que son muy distintos de la forma en que se diseñan los incentivos en el mercado.

La ventaja comparativa de la empresa con relación a los mercados se debe en parte a su capacidad única de utilizar incentivos de baja potencia combinados con restricciones. Esto explica por qué llevar el mercado al interior de la empresa es una idea tan mal orientada, algo que no comprendí en Ahlström y en lo que los defensores de los incentivos de mercado en las empresas parecen fallar hoy.

## Apéndice

### *A1. El Laboratorio General Multitarea*

En Holmström y Milgrom (1987) estudiamos un proceso browniano multidimensional en el que el agente puede elegir las tasas de deriva de forma independiente, a un costo que depende de las tasas elegidas. Este modelo presenta una solución óptima que es lineal en las diferentes dimensiones del proceso browniano, y al igual que en el caso unidimensional, los coeficientes óptimos pueden ser resueltos mediante un modelo estático en el cual el agente tiene varias tareas por realizar, cada una correspondiente a una de las dimensiones del proceso browniano. En su forma más general, el modelo multitarea estático tiene los siguientes elementos:

- El agente elige "insumos"  $e = (e_1, \dots, e_n)$ .
- Hay  $m$  medidas de rendimiento  $x_i = k_i(e) + \varepsilon_i$ ,  $i = 1, \dots, m$ . Los errores de medición siguen una distribución normal conjunta con el valor medio 0 y matriz de varianzas-covarianzas  $\Sigma$ . Las funciones de "producción"  $k_i$  determinan cómo las elecciones del agente se mapean con la media de la medida de desempeño correspondiente.
- La función de beneficio del principal es  $B(e)$  y la función de costo del agente es  $C(e)$ .
- El esquema de incentivos óptimo es lineal:  $s(x) = \sum_i \alpha_i x_i + \beta$ , con tasas de comisión  $\alpha_i$  y salario  $\beta$ .

El comportamiento del agente se caracteriza por las condiciones de primer orden

$$\sum_i \alpha_i \partial k_i(e) / \partial e_j = \partial C(e) / \partial e_j \text{ para todo } j.$$

Se aprecia que el modelo de desalineación es un caso especial de este modelo. El modelo también abarca los casos en que un agente neutral al riesgo observa una señal aleatoria  $\theta$  antes de actuar como en Baker (1992).

Ver Holmström y Milgrom (1991) para más detalles sobre la solución y variantes.

## Referencias

- Anderson, Erin. 1985. "The Salesperson as Outside Agent or Employee: A Transactions Cost Analysis," *Marketing Science* 4 (3): 234–54.
- Anderson, Erin and David C. Schmittlein. 1984. "Integration of the Sales Force: An Empirical Examination," *RAND Journal of Economics* 15 (3): 385–95.
- Baiman, Stanley, and Joel S. Demski. 1980. "Economically Optimal Performance Evaluation and Control Systems," *Journal of Accounting Research* 18 (Supplement): 184–220.
- Baker, George P., 1992. "Incentive Contracts and Performance Measurement," *Journal of Political Economy* 100 (3): 598–614.
- Baker, George. 2002. "Distortion and Risk in Optimal Incentive Contracts," *Journal of Human Resources* 37 (4): 728–51.
- Baker, George, Robert Gibbons, and Kevin J. Murphy. 1994. "Subjective Performance Measures in Optimal Incentive Contracts," *Quarterly Journal of Economics* 109 (4): 1125–56.
- Baker, George, Robert Gibbons, and Kevin J. Murphy. 1999. "Informal Authority in Organizations," *Journal of Law, Economics, and Organization* 15 (1): 56–73.
- Baker, George, and Bjorn Jorgensen. 2003. "Volatility, Noise and Incentives." <http://www.bristol.ac.uk/media-library/sites/cmpo/migrated/documents/baker.pdf>
- Bebchuk, Lucian A., and Jesse M. Fried. 2004. *Pay without Performance: The Unfulfilled Promise of Executive Compensation*. Cambridge, MA: Harvard University Press.
- Carroll, Gabriel. 2015. "Robustness and Linear Contracts," *American Economic Review* 105 (2): 536–63.
- Chaigneau, Pierre, Alex Edmans, and Daniel Gottlieb. 2014. "The Informativeness Principle Under Limited Liability," *National Bureau of Economic Research Working Paper* 20456.
- Chassang, Sylvain. 2013. "Calibrated Incentive Contracts," *Econometrica* 81 (5): 1935–71.
- DeMarzo, Peter M., and Yuliy Sannikov. 2006. "Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model," *Journal of Finance* 61 (6): 2681–2724.
- Dewatripont, Mathias, Ian Jewitt, and Jean Tirole. 1999. "The Economics of Career Concerns, Part II: Application to Missions and Accountability of Government Agencies," *Review of Economic Studies* 66 (1): 199–217.
- Diamond, Peter. 1998. "Managerial Incentives: On the Near Linearity of Optimal Compensation," *Journal of Political Economy* 106 (5): 931–57.
- Edmans, Alex, and Xavier Gabaix. 2011. "Tractability in Incentive Contracting," *Review of Financial Studies* 24 (9): 2865–94.
- Edmans, Alex, Xavier Gabaix, Tomasz Sadzik, and Yuliy Sannikov. 2012. "Dynamic CEO Compensation," *Journal of Finance* 67 (5): 1603–47.
- Garicano, Luis, and Luis Rayo. 2016. "Why Organizations Fail: Models and Cases," *Journal of Economic Literature* 54 (1): 137–92.
- Gjesdal, Frøystein. 1982. "Information and Incentives: The Agency Information Problem," *Review of Economic Studies* 49 (3): 373–90.

- Grossman, Sanford J., and Oliver D. Hart. 1983. "An Analysis of the Principal-Agent Problem," *Econometrica* 51 (1): 7–45.
- Grossman, Sanford J., and Oliver D. Hart. 1986. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy* 94 (4): 691–719.
- Healy, Paul M. 1985. "The Effect of Bonus Schemes on Accounting Decisions," *Journal of Accounting and Economics* 7 (1–3): 85–107.
- Hellwig, Martin F., and Klaus M. Schmidt. 2002. "Discrete-Time Approximations of the Holmström-Milgrom Brownian-Motion Model of Intertemporal Incentive Provision," *Econometrica* 70 (6): 2225–64.
- Holmström, Bengt. 1977. "On Incentives and Control in Organizations," PhD diss., Stanford University.
- Holmström, Bengt. 1979. "Moral Hazard and Observability," *Bell Journal of Economics* 10 (1): 74–91.
- Holmström, Bengt. 1982. "Moral Hazard in Teams," *Bell Journal of Economics* 13 (2): 324–40.
- Holmström, Bengt. 1989. "Agency Costs and Innovation," *Journal of Economic Behavior and Organization* 12 (3): 305–27.
- Holmström, Bengt. 1999a. "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies* 66 (1): 169–82.
- Holmström, Bengt. 1999b. "The Firm as a Subeconomy," *Journal of Law, Economics, and Organization* 15 (1): 74–102.
- Holmström, Bengt, and Paul Milgrom. 1987. "Aggregation and Linearity in the Provision of Incentives," *Econometrica* 55 (2): 303–28.
- Holmström, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization* 7: 24–52.
- Holmström, Bengt, and Paul Milgrom. 1994. "The Firm as an Incentive System," *American Economic Review* 84 (4): 972–91.
- Holmström, Bengt, and Joan Ricart i Costa. 1986. "Managerial Incentives and Capital Management," *Quarterly Journal of Economics* 101 (4): 835–60.
- Hurwicz, Leonid, and Leonard Shapiro. 1978. "Incentive Structures Maximizing Residual Gain under Incomplete Information," *Bell Journal of Economics* 9 (1): 180–91.
- Innes, Robert D. 1990. "Limited Liability and Incentive Contracting with Ex-ante Action Choices," *Journal of Economic Theory* 52 (1): 45–67.
- Kerr, Steven. 1975. "On the Folly of Rewarding A, While Hoping for B," *Academy of Management Journal* 18 (4): 769–83.
- Kim, Son Ku. 1995. "Efficiency of an Information System in an Agency Model," *Econometrica* 63 (1): 89–102.
- Lazear, Edward P. 1989. "Pay Inequality and Industrial Politics," *Journal of Political Economy* 97 (3): 561–80.
- Lazear, Edward P. 2000. "Performance Pay and Productivity," *American Economic Review* 90 (5): 1346–61.



- Lazear, Edward P., and Sherwin Rosen. 1981. "Rank Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy* 89 (5): 841–64.
- Levin, Jonathan. 2003. "Relational Incentive Contracts," *American Economic Review* 93 (3): 835–57.
- MacLeod, W. Bentley, and James M. Malcomson. 1988. "Reputation and Hierarchy in Dynamic Models of Employment," *Journal of Political Economy* 96 (4): 832–54.
- Milgrom, Paul R. 1981. "Good News and Bad News: Representation Theorems and Applications," *Bell Journal of Economics* 12 (2): 380–91.
- Milgrom, Paul R., and John Roberts. 1988. "An Economic Approach to Influence Activities in Organizations," *American Journal of Sociology* 94: S154–S179.
- Milgrom, Paul R., and John Roberts. 1990. "The Economics of Modern Manufacturing: Technology, Strategy, and Organization." *American Economic Review* 80 (3): 511–28.
- Milgrom, Paul R. and John Roberts. 1995. "Complementarities and Fit: Strategy, Structure and Organizational Change in Manufacturing," *Journal of Accounting and Economics* 19 (2–3): 179–208.
- Mirrlees, J. A. (1975) 1999. "The Theory of Moral Hazard and Unobservable Behaviour: Part I," *Review of Economic Studies* 66 (1): 3–21. Paper completed 1975, published 1999.
- Murphy, Kevin J. 1999. "Executive Compensation." In *Handbook Of Labor Economics*, Vol.3B, edited by Orly Ashenfelter and David Card, 2485–2563. New York: Elsevier.
- Neal, Derek. 2011. "Design of Performance Pay in Education." In *Handbook of Economics of Education*, Vol. 4, edited by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, 495–550. Amsterdam: Elsevier.
- Oyer, Paul. 1998. "Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality," *Quarterly Journal of Economics* 113 (1): 149–85.
- Prendergast, Canice. 1993. "A Theory of 'Yes Men'," *American Economic Review* 83 (4): 757–70.
- Prendergast, Canice. 2002. "The Tenuous Trade-Off between Risk and Incentives," *Journal of Political Economy* 110 (5): 1071–1102.
- Rogerson, William P. 1985. "The First-Order Approach to Principal–Agent Problems," *Econometrica* 53 (6): 1357–67.
- Ross, Stephen A. 1973. "The Economic Theory of Agency: The Principal's Problem," *American Economic Review* 63 (2): 134–39.
- Sannikov, Yuliy. 2008. "A Continuous-Time Version of the Principal-Agent Problem," *Review of Economic Studies* 75 (3): 957–84.
- Shavell, Steven. 1979. "Risk Sharing and Incentives in the Principal Agent Relationship," *Bell Journal of Economics* 10 (1): 55–73.
- Spence, Michael, and Richard Zeckhauser. 1971. "Insurance, Information, and Individual Action," *American Economic Review* 61 (2): 380–87.
- Stiglitz, Joseph E. 1975. "Incentives, Risk, and Information: Notes Towards a Theory of Hierarchy," *Bell Journal of Economics* 6 (2): 552–79.
- Tayan, Brian. 2016. "The Wells Fargo Cross-Selling Scandal," <https://www.gsb.stanford.edu/facultyresearch/publications/wells-fargo-cross-selling-scandal>

Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting*. New York: Free Press.

Wilson, Robert. 1969. "The Structure of Incentives for Decentralization under Uncertainty," In *La Decision: Agrégation et Dynamique des Ordres de Préférence*, edited by Georges-Théodule Guilhaud, 287–307. Paris: Centre National de la Recherche Scientifique.