

¿Cómo y a quién incentivar?*

Mecanismos Intrapersonales e Interpersonales

VITTORIO PELLIGRA

“Una simple posibilidad es que los modelos económicos que ignoran la psicología social son descripciones incompletas de los incentivos en las organizaciones. Una posibilidad más perturbadora es que las prácticas gerenciales basadas en modelos económicos pueden perjudicar (o incluso destruir) realidades no económicas como las motivaciones intrínsecas y las relaciones sociales.”

R. Gibbons (1998, p. 130)

1. Introducción

Aproximadamente en la última década, se ha aceptado ampliamente que “los incentivos son la esencia de la economía” (Predergast, 1999, p.7). Ha aparecido una gran cantidad de literatura que analiza todos los aspectos relevantes acerca de cómo funcionan los incentivos y cómo los sistemas de provisión de incentivos deben ser diseñados para promover la eficiencia. Cerca de la totalidad de esa corriente de estudios se basa en la suposición que los agentes son maximizadores clásicos de utilidad individual. Por “maximizadores clásicos de utilidad individual” me refiero a un agente que actúa con el deseo de alcanzar lo preferido entre los resultados que sus acciones puedan generar, y como corolario de esa suposición, las recompensas materiales juegan un papel dominante en la formación del orden de preferencias del agente. A pesar de que tal presunción es más comúnmente

utilizada cuando se reproducen hechos económicos, un creciente cuerpo de evidencia ha empezado a reunirse en años recientes, poniendo en duda su precisión descriptiva, tanto en situaciones paramétricas como estratégicas. Es cada vez más claro que al enfrentarse con cierta clase de problemas de toma de decisiones, las personas no se ajustan sistemáticamente a las predicciones formuladas por la Teoría de Utilidad Esperada y por la Teoría de los Juegos. Sin embargo es posible racionalizar y explicar dichas anomalías al considerar una versión extendida de la racionalidad, donde la maximización de la utilidad material ya no es el único motor para el accionar, y los agentes no sólo están interesados en los resultados de sus elecciones, sino también en el proceso que lleva a aquellos resultados. Aquí exploraré formas de completar dicha descripción de agencia humana. Algunas de las enmiendas requeridas son compatibles con el marco básico de la Teoría de Elección Racional (Rational Choice Theory), mientras que otras requieren el desarrollo de nuevos conceptos. En ambos casos, la incorporación en la estructura motivacional de agentes económicos de aquellos elementos basados en lo psicológico, mejora el poder predictivo de la teoría económica, permitiéndonos comprender con mayor profundidad el funcionamiento de los incentivos en el proceso de toma de decisiones, y sus diversas consecuencias

importantes en el área normativa del diseño institucional.

El centro de este artículo será, por lo tanto, el funcionamiento de los incentivos (Cómo) y el modelo de agente supuestamente sujeto a tales incentivos (Quién).

El artículo está organizado de la siguiente forma: en primer lugar, se describen los principios básicos de la teoría clásica de la agencia (§ 2); en segundo lugar se presentan ejemplos que demuestran la importancia de los mecanismos motivacionales intrapersonales (§ 3); en tercer lugar se analizan críticamente dos de las principales hipótesis de la teoría clásica (§ 4); en cuarto lugar se discute el rol de los mecanismos interpersonales (§ 5), y se provee un marco teórico basado en la aprobación social, la reciprocidad y la confianza, el cual justifica la evidencia empírica mencionada anteriormente (§ 6). Luego se establecen las implicancias para el diseño institucional (§ 7). Las conclusiones son el final del artículo (§ 8).

2. Cómo deben funcionar los incentivos... en teoría

La teoría de la agencia presupone dos tipos de sujetos, el principal y el agente. A menudo se piensa que los principales tienen algún interés que no puede perseguirse sin la participación de el/los agente/s. Dos hechos principales caracterizan la relación principal-agente: primero, existe un conflicto de intereses, y segundo, las acciones o características del agente sólo pueden ser observadas en forma imperfecta por el principal. Dicho en forma más simple, tal relación sería la existente entre empleador y empleado. El empleador tiene como objetivo maximizar las ganancias, las cuales dependen positivamente del esfuerzo del empleado. El empleado es adverso al esfuerzo y el nivel de esfuerzo realmente llevado a cabo no puede ser observado directamente por el principal. Un salario, el cual constituye un costo para el empleador y una fuente de utilidad para el empleado, debe ser provisto al empleado para incentivarlo a desempeñar cierto nivel de esfuerzo. En

este sentido, un contrato es una herramienta diseñada para alinear los objetivos conflictivos del empleador y del empleado. El salario provisto por el empleador debe satisfacer ciertos requisitos: desde el punto de vista del empleador, debe ser al menos tan alto como su utilidad de reserva (restricción de participación); y considerando esto, el empleado realizará un esfuerzo que maximice su utilidad neta: la diferencia entre la utilidad derivada del salario y la desutilidad derivada del trabajo debe ser positiva (restricción de compatibilidad de incentivo).

Existen tres postulados subyacentes a dicha teoría clásica según los cuales:

1. Cuánto mayor sea el salario, mayor será el esfuerzo ejercido;
2. Las personas solo están interesadas en los resultados a los que llevan sus acciones;
3. Dada la asimetría en la estructura de la información, cada vez que sea posible, el agente se comportará en forma oportunista.

La mayoría de los desarrollos recientes en la teoría de la agencia apuntan a hallar mecanismos de compensación óptimos capaces de reducir el riesgo de oportunismo, mientras que a su vez diseñan un contrato lo suficientemente atractivo en lo contextual, de modo que sea aceptado por el agente. Algunos de los aportes, por ejemplo, buscan crear instrumentos para superar la observabilidad imperfecta del esfuerzo del agente, vinculándolo a diferentes elementos observables. Otros introducen consideraciones dinámicas para extender la relación a lo largo del tiempo. Dicha medida permite a los teóricos explicar el rol de la reputación. Otros se centran en los problemas que surgen del disfrute libre (*free-riding*) en el trabajo en equipo y demás cuestiones¹. A pesar del debate activo que se lleva a cabo en este área parece, sin embargo, que los tres supuestos mencionados anteriormente en general no han sido puestos en duda.

A continuación discutiré algunos fenómenos que no concuerdan con dichos supuestos. En particular, en la siguiente sección describiré patrones de comportamiento que están en disconformidad con la rela-

ción positiva afirmada por la suposición 1), entre incentivos materiales y desempeño, y que refutan la orientación consecuencialista de la teoría clásica de la agencia, como se describe en la suposición 2). Este último hecho demostrará la naturaleza simplista de la suposición 3).

3. Cómo parecen funcionar los incentivos... en la práctica: motivaciones intrapersonales

A menudo se cree que la hipótesis 1) es una ley general del comportamiento humano. Está tan bien fundamentada y es tan general que se ha transformado casi en un axioma. Mientras que por un lado, es cierto que la suposición ha recibido cierto grado de validación empírica, muchos, de hecho, han hallado una correlación estricta, en el conjunto, entre el incremento en el salario y el incremento en la productividad, por otro lado, tal respaldo puede interpretarse de varias maneras. En primer lugar, debe tenerse en cuenta que la correlación entre remuneración y productividad, la cual es el principal hallazgo empírico, puede tener una explicación dual: primero, cuanto más se le paga a un individuo, mejor será su desempeño, según el supuesto. Segundo, cuanto más se ofrece un empleo, mayor será la probabilidad de atraer trabajadores calificados con una productividad más elevada. Por consiguiente, la correlación observada puede explicarse tanto a través del “asunto-argumento del incentivo” como del “argumento de selección”. Más adelante presentaré ejemplos que ilustran estos dos argumentos. Comenzaré con el argumento de selección.

Este razonamiento mantiene que al incrementar la remuneración para una determinada tarea uno atrae a sujetos que están mejor capacitados para dicha tarea. Consideremos los siguientes ejemplos. En un estudio de gran influencia acerca de la entrega de donaciones (*gift-giving*), el sociólogo Richard Titmuss (1970), halló que, a pesar de ser voluntario, el sistema de donación de sangre adoptado en Inglaterra era más eficiente (en lo que respecta al volu-

men, calidad y disponibilidad temporal de la sangre recibida) si se lo comparaba con el sistema remunerado utilizado en los Estados Unidos en aquella época. Si se paga por donar sangre se llega a una menor calidad y cantidad del suministro de sangre. En un intento de incrementar el suministro de sangre, los estadounidenses permitieron a los bancos de sangre comerciales abonar a los donantes por su sangre. En contraposición a lo que se hubiera esperado el resultado fue desastroso en todas las dimensiones.

En el estudio llevado a cabo por Barkema (1995) se comparan dos grupos de gerentes sujetos a dos regímenes diferentes de supervisión. El Grupo A tiene un mayor grado de libertad, mientras que el Grupo B es controlado estrictamente. La idea subyacente es que a medida que el monitoreo se torna más riguroso será más fácil observar el esfuerzo de cada agente y recompensarlo en consecuencia. Esa correlación estricta entre el esfuerzo y la recompensa debería ejercer un efecto positivo en a nivel del esfuerzo en sí mismo. Sin embargo, Barkema presenta un resultado desconcertante, ya que de hecho, el efecto de dichos regímenes diferentes es que el Grupo B, aquel supervisado en forma más estricta, tiene un rendimiento menor que el Grupo A que no es monitoreado.

Un tercer ejemplo destaca el mismo efecto contraproducente. Gneezy y Rustichini (2000) llevaron a cabo un estudio utilizando padres que enviaban a sus hijos a un jardín de infantes israelí. Analizaron durante 20 semanas cómo reaccionaban los padres a la introducción de una multa para los que se retrasaban en pasar a buscar a sus hijos causando, en esta forma, problemas al personal. La multa tenía como fin, por medio de la imposición de un costo adicional, reducir el número de padres que llegaban tarde, el cual, por el contrario aumentó en forma sorprendente. El número elevado se mantuvo estable aún luego de la eliminación de la multa.

Estos ejemplos realzan la naturaleza problemática tanto del “argumento de selección” como del “asunto-argumento del

incentivo”. Como veremos Titmuss limita su explicación a la diferencia en la clase de razones que impulsan a *diferentes* agentes (es decir que diferentes personas reaccionan de manera diferente al mismo incentivo monetario); los otros dos ejemplos van más allá, centrándose en la diferencia en la clase de razones que movilizan al *mismo* agente (es decir, el desempeño del mismo agente puede verse afectado en forma positiva o negativa por el mismo incentivo material).

Este último efecto puede explicarse considerando el llamado “efecto de exclusión” (Frey, 1997; Frey – Oberholzer-Gee, 1997). En ciertos casos la predisposición de los sujetos a llevar a cabo una determinada acción disminuye (en lugar de aumentar, como sugeriría la intuición) con la perspectiva de una recompensa monetaria o material. Titmuss acentuó la importancia de concientizarse de que mientras existen personas que desarrollan una cierta tarea con el objetivo de obtener alguna forma de recompensa material (motivación extrínseca) otros pueden tener razones intrínsecas para desarrollar ciertas acciones, ya que al introducir la recompensa monetaria se tenderá a seleccionar solo a los primeros sujetos. La teoría de la exclusión afirma que la *misma* persona puede tener tanto motivaciones extrínsecas como intrínsecas, y que cuando uno trata de incrementar, a través de recompensas materiales, la predisposición de los sujetos a desarrollar ciertos tipos de acciones regidas por motivaciones intrínsecas, la motivación subyacente se ve transformada de intrínseca a extrínseca, y el resultado general es una disminución en la predisposición del agente a llevar a cabo esa clase de acciones.

4. Una crítica del “Argumento de selección” y del “Asunto-argumento de los incentivos”

Los ejemplos delineados anteriormente acentúan cómo los dos aspectos que subyacen al supuesto 1), al cual he denominado “el argumento de selección” y el “asunto-argumento de los incentivos”, poseen en ciertas circunstancias, una validez empírica limitada.

Consideremos la explicación de Titmuss. Su argumento se basa en los diferentes tipos de motivaciones que se hallan en una misma acción (donación de sangre). En el caso del donante voluntario, el motivo es altruista y en pro de los demás, se basa en razones intrínsecas; mientras que en el caso del donante remunerado, el motivo es puramente egoísta, se basa en motivaciones extrínsecas. Según Titmuss esto lleva a una autoselección de donantes potenciales, por un lado aquellos motivados intrínsecamente y por el otro, individuos con un mayor interés en la recompensa material, extrínseca; personas con una mayor tendencia a ser oportunistas que pueden, por ejemplo, ser menos sinceras acerca del riesgo de hepatitis B. Se observó que los donantes potenciales con motivaciones intrínsecas fueron desplazados por la introducción de una recompensa monetaria en los bancos de sangre comerciales estadounidenses. Esta autoselección afectó marcadamente la calidad de la sangre de hecho provista.

Mientras que Titmuss acentúa el riesgo de la selección adversa implícito en el uso de los incentivos materiales, los otros dos ejemplos demuestran cómo la introducción de una recompensa monetaria puede desalentar a las *mismas* personas a desarrollar la acción en particular que el incentivo debía fomentar. Las razones detrás de dicho fenómeno son varias; entre ellas son en especial relevantes aquéllas relacionadas con la autodeterminación y la autoestima del sujeto (Frey, 1997a,ch.I). La imposición de un sistema externo de incentivos materiales produce en los sujetos la impresión de ser controlados y de perder el control de la situación (Rotter,1966), de modo que el *locus* de motivación se traslada de lo interno a lo externo. Una intervención externa, en forma similar, puede transmitir el mensaje de que la responsabilidad individual del sujeto (y por consiguiente también el mérito potencial) no se reconoce y que la motivación intrínseca se rechaza. “A la persona intrínsecamente motivada se le niega la oportunidad de desplegar su propio interés y participación en una actividad cuando alguien más ofrece una recompensa” (Frey

1997^a, p.47). Como resultado de la subestimación de la responsabilidad, el sujeto experimenta una disminución de su autoestima, lo cual, como consecuencia, reduce su predisposición para poner en práctica una determinada acción. Más aún, la manera en que el sujeto percibe la intervención externa juega un papel crucial a la hora de determinar el efecto de exclusión o de inclusión. De hecho, tal intervención puede considerarse como una forma de controlar o de avalar el comportamiento de los sujetos. En el segundo caso se observa un fortalecimiento de las actitudes en pro de los demás de los sujetos (inclusión), en el primer caso, debido a una disminuida autodeterminación y autoestima, se observa su debilitamiento (exclusión).

Los ejemplos presentados en las secciones anteriores han sido a menudo analizados según el efecto de exclusión. Mientras que esta interpretación explica los denominados costos encubiertos de las recompensas en términos de la autoestima y de las intervenciones de apoyo o de control, esta estrategia de análisis no logra explicar otra clase de anomalías de comportamiento que se relacionan con el postulado 2) y que se refieren a los aspectos relacionales a la estructura motivacional del agente.

En la siguiente sección trataré algunos ejemplos de dicho comportamiento anómalo y luego intentaré proporcionar algunos elementos para construir un marco unificador con el que ambas clases de violaciones podrían explicarse.

5. Cómo los incentivos parecen funcionar... en la práctica: motivaciones interpersonales

Los ejemplos que he discutido hasta ahora demuestran cómo los incentivos pueden ser ineficientes en lograr su objetivo debido a que la coexistencia de motivaciones intrínsecas y extrínsecas a menudo se ignora y los sistemas de incentivo se diseñan como si sólo existieran motivaciones extrínsecas.

Presentaré ejemplos que señalan lo inadecuado de un modelo consecuencialista

de la agencia que presupone agentes exclusivamente interesados en el resultado de sus acciones.

Utilizando el llamado “juego de intercambio de dones”, Fehr y Gächter (1997) demuestran cómo, en contraposición a la predicción económica estándar, la participación en relaciones de intercambio reguladas por contratos no completamente ejecutorios podrían ser eficientes en términos de la maximización de la utilidad. El juego está enmarcado como una relación de trabajo, pero en la lógica subyacente puede aplicarse a un rango más amplio de relaciones contractuales. Se consideran dos tratamientos diferentes, el juego de dos etapas y una versión de tres etapas del mismo juego. La primera etapa es igual en ambos juegos y consiste en una oferta anunciada por seis principales que establece el salario ofrecido y el esfuerzo requerido. Cada principal ofrece sólo un contrato por período. Todas las ofertas son comunicadas a los ocho agentes en una sala diferente. Luego se le pide a los agentes que realicen una elección (aceptar/no aceptar) entre las ofertas a completar en la primera etapa. En la segunda etapa, los agentes deben determinar un nivel de esfuerzo asociado, de una forma predeterminada, a un cierto costo que también conocen los principales. Los esfuerzos y los costos consecuentes, se eligen en forma privada y son comunicados por un experimentador al principal elegido, quien no conoce la identidad del agente.

Este tipo de relación entre agentes y principal ha sido diseñada de esta manera para reflejar el estilo típicamente incompleto de la mayoría de los contratos de trabajo, lo cual implica un gran grado de discreción en la determinación del esfuerzo por parte del agente y una habilidad relativamente limitada del principal para hacer cumplir el nivel deseado; sólo el nivel mínimo puede de hecho hacerse cumplir de forma eficiente.

Luego de tomarse decisiones sobre el salario y el esfuerzo, se calculan los beneficios y se los asigna a los jugadores: el beneficio del principal (ganancia) se ve positivamente afectado por el esfuerzo de los agen-

tes y negativamente por el salario a pagar, mientras que el beneficio del agente (utilidad) es proporcional al salario y se ve negativamente afectado por el nivel de esfuerzo. Si el salario ofrecido es mayor al valor base y el esfuerzo expresado es igual al nivel mínimo de cumplimiento, el principal se enfrenta al riesgo de una gran pérdida. Es por este motivo que la inducción regresiva (backward induction) sugiere que el principal ofrecería solo el salario mínimo, y el agente ejercería el nivel mínimo de esfuerzo.

El tratamiento de las tres etapas es similar al de las dos etapas; la tercera etapa adicional, le otorga al principal la posibilidad de una reciprocidad (positiva o negativa), es decir, la oportunidad de recompensar o penalizar a los agentes. Los principales pueden elegir recompensar o penalizar, bajo su propio costo, las acciones de los agentes incrementando o disminuyendo los beneficios finales de los agentes.

En ambos tratamientos la predicción teórica del juego es el equilibrio subjuego perfecto de Nash, el cual involucra el más bajo nivel de esfuerzo y el más bajo nivel de salario. El equilibrio es el mismo en ambas versiones ya que las penas y recompensas se consideran en términos teóricos del juego, "amenazas no verosímiles" y por lo tanto no deberían afectar las elecciones de estrategia. A pesar de estas predicciones los resultados experimentales son diferentes en lo que concierne al nivel deseado de esfuerzo, los salarios ofrecidos y el nivel real de esfuerzo. El nivel deseado de esfuerzo es en promedio 0.70 y 0.72, respectivamente en el juego de dos y tres etapas, algo muy superior al nivel de 0.1 teóricamente esperado. Estos niveles son constantes a lo largo de todos los períodos sin ninguna señal de convergencia hacia el equilibrio. Los niveles de esfuerzo obtenidos son mucho más altos de lo que se habría esperado: un promedio de 0.44 y 0.63 en el juego de dos y tres etapas respectivamente, y se correlacionan en forma positiva con la remuneración ofrecida. El esfuerzo real es en promedio más alto en el juego de tres etapas. Esto demuestra como una "amenaza no verosímil" puede afectar los comportamientos en forma significativa. Estos hallaz-

gos son coherentes con la firme creencia de Herbert Simon (1991) según la cual "en la mayoría de las organizaciones, los empleados contribuyen bastante más al logro de metas que lo mínimo que podría lograrse con una supervisión más estricta" (p. 31-32).

Un quinto ejemplo destacará la importancia de tener potenciales incentivos materiales o desincentivos y no utilizarlos. Fehr y List (2002) observaron el comportamiento de una muestra de presidentes de compañías (*Chief Executive Officers - CEOs*) en varias formas de juegos de inversión. En particular consideraron dos variantes del juego, una donde el primer jugador recibe una cierta cantidad de dinero y debe decidir qué porcentaje de los recursos enviar al segundo jugador, o no enviarle nada. Si envía una cantidad positiva esta será triplicada y otorgada al segundo jugador quien podrá, a su vez, decidir cuánto enviar de vuelta, o no enviar nada. La segunda variante del juego es similar, excepto por el hecho de que el primer jugador puede implementar una sanción si lo que recibe del segundo jugador es menos de lo esperado. La solución teórica del juego a ambas variantes es que el primer jugador no envíe nada al segundo jugador, y que éste último tampoco envíe nada al primer jugador. No sólo la mayoría de los individuos bajo observación decidió invertir y enviar algún porcentaje de dinero, sino que el resultado más sorprendente presentado por los dos autores es que dicha cantidad aumentó cuando la sanción estaba disponible pero, de hecho, no fue utilizada. Este resultado es interpretado por Fehr y List como una señal de que: "la *disponibilidad* de la amenaza de sanción puede ser bastante productiva (. . .) Si los principales se abstienen voluntariamente de emplear la amenaza de penalizar cuando la misma está disponible, los agentes exhiben significativamente mayor honradez que si la amenaza de castigo no estuviera disponible. Por consiguiente, si los agentes no se enfrentan a una amenaza de castigo, el mero hecho de que el principal podría haber utilizado la opción de penalizar afecta la honradez del agente en forma positiva" (2002, p.2). Blair y Stout informan sobre

resultados similares (2000) con respecto a las leyes corporativas y su cumplimiento en la jurisprudencia americana.

Siguiendo la misma línea, si podemos comparar un “juego de recompensa a la confianza” (rewarding trust game) (Bolle, 1995) y un “juego del dictador” (dictator game) (Bohnet, Frey, 1999) desde la perspectiva del segundo jugador, vemos que se enfrenta exactamente al mismo juego. En el juego de recompensa a la confianza debe decidir cómo dividir un cierto monto de dinero entre él/ella y el primer jugador, y se le pide que haga exactamente lo mismo en el juego del dictador. ¿Por qué entonces observamos que en el juego del dictador el 25% de las elecciones consisten en una proporción igual, mientras que la misma elección es preferida por la mayoría de los segundos jugadores en el juego de recompensa a la confianza? La única diferencia en la estructura del juego yace en que en el juego de recompensa a la confianza, el jugador uno debe primero decidir si confía en el jugador dos o no. Por lo tanto, si el jugador dos tiene la oportunidad de tomar una decisión se debe a que el jugador uno ha confiado en él/ella. La diferencia entre los dos resultados radica en que las personas (los jugadores dos) no sólo se preocupan por los resultados, sino también, en este caso, por lo que el jugador uno podría haber hecho y no hizo.

Pelligra (2003) presenta los resultados de un experimento llevado a cabo en 51 individuos donde se le pidió a los jugadores participar en un juego de inversión y un juego del dictador. Desde la perspectiva del segundo jugador el juego es exactamente el mismo con la única (teóricamente no significativa) diferencia que en el juego de inversión hay una rama del juego a la que no se llega, una experiencia pasada (*bygone*). Aunque los dos juegos son idénticos desde el punto de vista teórico, es notable que en el juego de inversión el jugador otorga un promedio de 11 Euros, mientras que en el juego del dictador la oferta promedio es de sólo 5 Euros.

Del estudio realizado por Falk, Fehr y Fishbacher (1998) emergen consideracio-

nes similares. Compararon un Juego de Ultimatum Limitado (Reduced Ultimatum Game) con un Juego de mejor oportunidad limitada (Reduced Best-Shot Game). En el primer juego se le pide al primer participante que divida su patrimonio de 10, y presente una oferta a su oponente. Si el oponente acepta, luego la oferta se confirma y el juego finaliza, de lo contrario, si la oferta es rechazada ningún jugador obtiene nada. Las ofertas posibles son sólo dos (es por eso que esta versión del juego se denomina “limitada”): 2 ó 5. En el segundo juego, el Juego de Mejor Oportunidad Limitada se observa la misma estructura pero las ofertas posibles son 2 u 8, con un consiguiente resultado simétrico igual a (2,8) y (8,2).

¿Debe considerarse la oferta de 2 de la misma forma en ambos juegos? Los individuos experimentales sugieren que no. Los autores descubrieron, de hecho, que la frecuencia de rechazo de la oferta de 2 en el Juego de mejor oportunidad limitada (Reduced Best-Shot Game) es significativamente inferior a aquella de la misma oferta en el Juego de Ultimatum Limitado (Reduced Ultimatum Game). ¿Cómo explicamos tal diferencia? El asumir que los participantes son altruistas o reacios a la injusticia no contribuye demasiado. Estas teorías, de hecho, debido a su naturaleza progresista, son incapaces de explicar dichos patrones de comportamiento.

Nuevamente, la estrategia alternativa consiste en asumir que a las personas les interesa no sólo el resultado sino también los resultados y las intenciones. Las intenciones pueden inferirse al estudiar la estrategia elegida no en forma aislada sino como parte de un conjunto total de estrategias, esto le otorga al agente la posibilidad de evaluar lo que el rival podría haber hecho y no hizo. En el Juego de Ultimatum Limitado (Reduced Ultimatum Game), en realidad, los primeros jugadores que ofrecieron 2, podrían haber optado por una división en partes iguales pero no lo hicieron. Como causa de esta comparación entre la elección real y las otras opciones disponibles, dicha elección se percibe como menos bondadosa

que la misma elección en el Juego de mejor oportunidad limitada (Reduced Best-Shot Game), y se ve penado por un mayor número de rechazos.

En los resultados experimentales presentados en Nelson (2002) surgen consideraciones similares acerca de la diferencia entre explicaciones basadas en cuestiones de reparto y teorías basadas en la receptividad del agente, por ejemplo, modelos de reciprocidad.

En resumen, puede decirse que el hecho de que la cuestión de la experiencia pasada (*bygones*) sea relevante significa que no podemos centrar nuestra atención exclusivamente en las consecuencias de las elecciones, como el único criterio que los individuos emplean para evaluar sus acciones potenciales, sino que también los diferentes procedimientos que llevan a los resultados deberían tener relevancia teórica.

6. Incentivos relacionales: reciprocidad y confianza

La evidencia que he presentado hasta ahora, debería haber clarificado cómo el conservar los postulados 1) y 2) implica, de alguna manera, una reducción del poder descriptivo y explicativo de la teoría de la agencia. Más aún, a partir de la crítica del postulado 2), se concluye que el comportamiento oportunista no es tan general como sugiere la hipótesis 3). En realidad depende de la estructura particular de interacción en que se encuentran los participantes y del comportamiento de *todos* los participantes. La información parece sugerir que una versión más satisfactoria de la teoría de la agencia, debería incorporar una explicación conductual de cómo funcionan los incentivos, y un modelo más realista de agente económico. En esta sección trataré dos de los principios que pueden contribuir a construir dicha teoría, en otras palabras, reciprocidad y confianza.

En los experimentos de negociación que he mencionado, el “juego de intercambio de dones” (“gift exchange game”) (Fehr – Gächter, 1997), el “juego de inversión”

(Fehr – List, 2002), y el “juego de recompensa a la confianza” (Bolle, 1995), todos resaltan que la gente tiende a comportarse amablemente con las personas que han sido amables con ellos, y de manera descortés con los que han sido desatentos. Estos comportamientos responden a la norma de reciprocidad positiva y negativa. Se ha demostrado tanto empírica como teóricamente (Rabin, 1993) cómo, bajo ciertas condiciones, dicha norma puede compensar el efecto de las compensaciones materiales en la toma de decisiones estratégicas. El efecto de reciprocidad puede llevar al individuo a actuar de una manera que parece contraria a su propio interés material. La idea de reciprocidad se basa finalmente en los efectos *conjuntos* de incentivos materiales y psicológicos. Esto significa que la motivación que provoca el comportamiento recíproco (positivo o negativo) se basa en incentivos materiales. La amabilidad percibida que suscita un comportamiento recíproco, es una medida del beneficio material que la elección de un agente le atribuye a otro jugador.

Siguiendo precisamente esta lógica, Rabin, en su modelo, formaliza la idea de forma que el participante A se siente motivado a ser recíproco cuando, a partir del accionar esperado de B, puede obtener un beneficio *material* mayor al equitativo (lo cual constituye una medida de distribución justa). Dicho incremento se relaciona con los beneficios materiales. Actuar en una forma que corresponda (ignora) dicha ganancia material esperada, resulta en una gratificación (pérdida) psicológica para A.

Algunos de los otros experimentos que he tratado, sin embargo muestran que otro principio del comportamiento puede participar en situaciones similares, es decir la confianza, en particular, en la forma de receptividad de confianza (Bacharach, Guerra, Zizzo, 2002; Pelligra, 2002a, 2002b, 2002c). La principal característica de la confianza en este significado en particular, tiene que ver con el hecho de que un acto explícito de confianza posee la peculiaridad de “inducir” o “provocar”, hasta cierto punto, una respuesta fidedigna. En este sen-

tido se dice que la confianza es receptiva o autogratificante. Supongamos que tenemos dos agentes, A y B. Según la concepción de la “confianza receptiva”, la honradez de B puede verse influenciada por la elección de una acción confiable por parte de A (como, por ejemplo, el jugador 1 le envía dinero al jugador 2 en un juego de inversión, u ofrece un salario mayor al mínimo en el juego de intercambio de dones). Esta clase de aliciente presupone la existencia de un mecanismo psicológico según el cual, la acción confiable de A, motiva a B a recompensar dicha confiabilidad, haciendo que se comporte honradamente aunque tal comportamiento implique cierto costo material. Denomino a este mecanismo psicológico “receptividad de la confianza”.

Mientras que la reciprocidad, como he mencionado, se basa en la acción conjunta de incentivos materiales y psicológicos, la receptividad de la confianza se basa exclusivamente en una motivación psicológica-moral. Es la acción (confiable) la que juega un rol en la motivación de la persona en la que se confía, y no las potenciales consecuencias beneficiosas para la riqueza material que pueda obtener. Al confiar en alguien uno manifiesta sus expectativas acerca de su comportamiento. Si uno conscientemente cumple (frustra) las expectativas, uno obtiene un incremento (disminución) en su gratificación psicológica. Dicho de otra forma, mientras que la reciprocidad en la teoría de Rabin consiste en el acto de proporcionar beneficios a aquellas personas que nos han beneficiado previamente y en *forma material*, la receptividad de la confianza consiste en proveer beneficios a personas que han demostrado que esperan que lo hagamos, y que se han expuesto a un riesgo si actuamos según nuestro propio interés material.

Esta diferencia se torna más clara cuando se considera una cierta clase de situaciones donde los comportamientos basados en la reciprocidad y la receptividad de la confianza no pueden percibirse como equivalentes, sino que llevan a predicciones divergentes.

Consideremos la lógica de la reciprocidad aplicada al juego de la confianza deta-

llado en la figura 1a que aparece debajo. Sabemos que la reciprocidad tiende a dar amabilidad a cambio de amabilidad y descortesía por descortesía. En este juego esto significa que si A espera que B juegue R, luego “si A juega R” B lo percibe como amable, y a su vez se siente motivado a jugar R. Consideremos ahora la instancia del juego de la confianza descrita en la figura 1b. ¿Qué resultado podría producir la lógica de la reciprocidad en este juego en particular? Ya que ahora, si A espera que B juegue R, “si A juega R” esto no beneficia a B, si B espera que A juegue R, como resultado, la lógica de reciprocidad sugeriría que B jugará L, sin pensarlo. ¿Qué se puede decir sobre la receptividad de la confianza? Si, por el contrario, consideramos un jugador B receptivo a la confianza, observaríamos que él elige R en ambas situaciones, a pesar de la diferencia en la gratificación material que obtendría por la decisión de A. He catalogado esta última forma particular del juego de confianza como el “Juego de Confianza Gratuita” a fin de enfatizar que en este juego cualquier instancia de comportamiento de colaboración, no puede ser vista como un intercambio, aunque demorado, de beneficios materiales.

Figura 1a.
El Juego de la Confianza

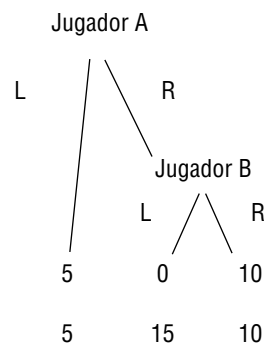
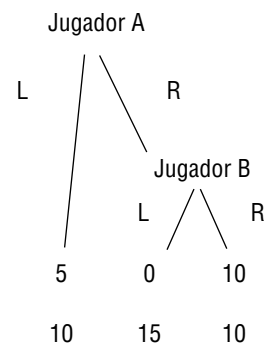


Figura 1b.
El Juego de la Confianza



Estos ejemplos amplían el panorama de la diferencia entre la receptividad a la confianza y la reciprocidad en la cuestión de los incentivos materiales/no materiales a la acción. En el Juego de Confianza Gratuita, de hecho, el participante B obtendría el mismo beneficio material, tanto de la elec-

ción confiada de A (seguida por la elección confiable de B) como de la elección desconfiada de A (L). De manera diferente al juego de confianza simple, la única forma de que B perciba una ganancia mayor a 10, en el Juego de Confianza Gratuita, es a través de una posición oportunista. Esto significa que si uno observa que un participante B juega R en la versión gratuita del juego de confianza, ese comportamiento no puede explicarse en términos de reciprocidad, aunque aún sea consistente con una explicación basada en la hipótesis de receptividad a la confianza. La relevancia empírica de tal patrón de comportamiento receptivo a la confianza se encuentra documentada en Bacharach, Guerra, Zizzo, (2002) y Pelligra (2002^a).

Ya debería verse con más claridad cómo la lógica de la receptividad a la confianza puede subsumir al efecto de exclusión. Una recompensa material, una remuneración por realizar una tarea que el agente estaba dispuesto a efectuar sobre la base de motivaciones intrínsecas, transmite una sensación de desconfianza que da lugar al oportunismo.

Como sabemos por el funcionamiento de la receptividad a la confianza, los agentes tienden a cumplir con las expectativas-manifiestas de los otros. En el caso de los efectos de exclusión, surge un conflicto entre las razones internas y externas para el accionar de los agentes. Tomemos el ejemplo de un trabajador que tiene un rendimiento pobre cuando se lo supervisa. Desilusionará a su principal, pero al mismo tiempo, reaccionará según su sentido de valor y autoestima, a un acto hostil. Mientras que la teoría motivacional de la atracción (crowding motivation theory) explica diferentes clases de comportamientos presumiendo la existencia de dos tipos diferentes de motivos para la acción (intrínsecos y extrínsecos), la receptividad a la confianza sugiere que los diversos efectos de los desincentivos dependen de la importancia relativa de las consecuencias sociales y psicológicas. Siguiendo la *Teoría de los Sentimientos Morales* de Smith, podemos hallar tres fuentes principales de la motivación: material, social y

psicológica. Las motivaciones materiales se relacionan a los resultados materiales de nuestras acciones; las motivaciones sociales se refieren al grado de aprobación social y reconocimiento que proviene de aquellos que observan nuestras acciones. La tercera fuente de motivación, la psicológica, tiene que ver con el grado en que consideramos que nuestras acciones son realmente merecedoras de tal reconocimiento o culpa. Deseamos, en realidad, no sólo ser elogiados sino ser dignos de elogio. En las propias palabras de Smith:

“El hombre naturalmente desea no sólo ser amado, sino ser amable; o ser aquello que es el objeto apropiado y natural del amor. Naturalmente teme, no sólo ser odiado, sino ser aborrecible; o ser objeto natural y apropiado del odio. Desea no sólo el elogio, sino ser digno de elogio; o ser aquello que, aunque no es elogiado por nadie, es sin embargo, el objeto natural y apropiado del elogio. Teme no sólo el reproche, sino ser digno de reproche; o ser aquello que, aunque no es reprochado por nadie es sin embargo el objeto natural y apropiado del reproche.”

(Smith, 1759/1976, sec.III.2.1)

7. Implicancias para la política y el diseño institucional

Mientras que la exclusión motivacional, a pesar de no ser estándar en la teoría de la agencia, puede fácilmente incorporarse en el modelo clásico de racionalidad económica (rational choice) (Frey, Oberholzer-Gee, 1997), la reciprocidad y la confianza, por otro lado, implican que los jugadores son perceptivos al comportamiento de los otros jugadores, es decir, que los beneficios son endógenos. Dicha característica posibilita reconciliar aquellos principios con el modelo estándar de la teoría de los juegos, la cual es esencialmente consecuencialista (Geanakoplos, Pearce, Stacchetti, 1989; Rabin, 1993).

Además de sus implicancias teóricas, las regularidades de comportamiento que he delineado, también poseen algunas impli-

cancias normativas, en especial en lo que respecta al diseño de esquemas de incentivos óptimos. Si las personas de hecho obtienen una utilidad psicológica de la autoestima, aprobación social, comportamiento recíproco y honradez, estos elementos deberían incorporarse en los sistemas de incentivo y utilizarse como importantes recursos organizacionales. A fin de evitar los efectos de exclusión, las recompensas materiales deben apuntar cuidadosamente a transmitir una sensación de respaldo en lugar de una sensación de control. Esta última puede tener un efecto adverso, reduciendo la motivación intrínseca de los individuos para llevar a cabo la misma acción que los incentivos deberían favorecer.

Más aún, ya que la reciprocidad y la confianza son elementos motivacionales activos, los esquemas de interacción deberían crearse dentro de organizaciones capaces de activar dichos elementos. Una supervisión demasiado estricta, por ejemplo, no daría lugar a la honradez intrínseca socialmente aprobada, con el riesgo de proyectar una sensación de desconfianza que puede aumentar el oportunismo y la falta de atención a las obligaciones (*shirking*), en lugar de reducirlo, como demuestra claramente el experimento de Barkema. Si se establecen incluso los detalles más específicos de un contrato, se puede restar espacio a acciones recíprocas y como se ha visto en el juego de intercambio de dones, se pueden obtener resultados inferiores de Pareto.

La reciprocidad y la confianza son normas que se hacen respetar a través de la presión social interpersonal. Como Fehr y Falk (2001) han establecido recientemente, es probable que tales normas produzcan una complementariedad estratégica entre las acciones de los agentes. Esto significa que la eficacia del motivo de la aprobación social depende del comportamiento del resto. Si otras personas son sensibles a la aprobación y desaprobación de sus pares, la acción de cada agente encontrará un gran incentivo psicológico en la búsqueda del elogio de los demás, de otro modo siempre predominará el interés material. Esto abre la posibilidad de que surjan múltiples equi-

librios medibles mediante Pareto. La transición de un equilibrio ineficiente a uno más eficiente dependería entonces de cómo los incentivos sociales funcionan dentro de la organización. Se puede entonces crear una “atmósfera” donde se promueva el deseo de reconocimiento y que finalmente favorezca que los agentes sean dignos de elogios, o al menos que los agentes busquen el reconocimiento social.

Tal como dijo Matthew Rabin: “La economía debería ocuparse no sólo de la distribución eficiente de bienes materiales, sino también de diseñar instituciones donde las personas estén conformes con la forma de interacción con los demás” (1993, p. 1283). Precisamente en este sentido, las motivaciones internas e intrínsecas de los agentes deberían considerarse como realidades económicas, así también como su búsqueda de una recompensa material. Se necesita un diseño cuidadoso para evitar el riesgo de producir un conflicto entre ambos tipos de motivaciones.

Las motivaciones intrínsecas, la confianza y la reciprocidad deben tenerse en cuenta como componentes relevantes, y en ocasiones cruciales, de cada organización. Si se ignora este punto podrían producirse efectos contraproducentes, con un resultante desperdicio de recursos y perjuicio a la eficiencia organizacional.

8. Conclusiones

En este artículo he discutido y cuestionado en forma crítica, los tres principales postulados de la teoría de la agencia. Respectivamente: a) a mayor salario mayor será el esfuerzo ejercido, b) las personas sólo están interesadas en los resultados de sus acciones, c) dada la asimetría de la estructura de la información, siempre que sea posible, el agente tendrá una actitud oportunista. De acuerdo a esta visión estándar los agentes deben juzgarse como “calculadores egocéntricos”, citando la expresión de Williamson (1985).

Mi posición alternativa sostiene que:

Debido a la interacción de motivos

intrínsecos y extrínsecos, puede ser posible que la utilización de premios materiales para incentivar actividades motivadas intrínsecamente, reduzca el rendimiento en tales actividades (exclusión motivacional); las personas son receptivas al comportamiento de los demás, por lo tanto el mismo resultado puede ser evaluado de manera diferente según las estrategias que lleven al mismo; por esta razón las personas tienden al oportunismo en una proporción bastante menor a la que sugeriría la teoría clásica. La confianza y la reciprocidad son los principios que explican los comportamientos anómalos observados.

Estos puntos tienen implicancias significativas para la actividad del diseño institucional. El anhelo por la aprobación social, la confianza y la reciprocidad, deben tenerse en cuenta como recursos organizacionales que deberían direccionarse cuidadosamente para evitar efectos contraproducentes, y para mejorar el desempeño general.

Bibliografía

Bacharach M., Guerra G., Zizzo D. (2001), "Is Trust Self-Fulfilling? An Experimental Study", mimeo, BREB, University of Oxford.

Barkema H., (1995), "Do Top Managers Work Harder When They Are Monitored?", *Kyklos*, 48: 19-42.

Blair M., Stout L. (2000), "Trust, Trustworthiness, and the Behavioral Foundations of Corporate Law", Working Paper, Georgetown University Law Centre.

Bohnet I. – Frey B., (1999), "Social Distance and Other-Regarding Behaviour in Dictator Game: Comment", *American Economic Review*, 89:335-39.

Bolle F., (1998), "Does Trust Pay?", Europa Universität Frankfurt (Oder), Working Paper.

Geneakoplos J. – Pearce D. – Stacchetti E., (1989), "Psychological Games and Sequential Rationality", *Game and Economic Behavior*, 1:60-79

Gibbons R., (1998), "Incentives in Organisations", *Journal of Economic Perspectives*, 12: 15-132

Gneezy U. – Rustichini A., (2000), "A Fine is a Price", *Journal of Legal Studies*, 29:1-17.

Falk A. - Fehr E. - Fischbacher U., (1998), "Intentions Matter", *mimeo*, University of Zurich.

Fehr E. - Gächter S., (1997), "How effective are Trust- and Reciprocity-Based Incentives?" in Ben-Ner A. – Putternam L. (Eds.), *Economics, Values and Organisation*, Cambridge, Cambridge University Press.

Fehr E., - List J., (2002), "The Hidden costs and Returns of Incentives – Trust and Trustworthiness among CEOs", Mimeo – University of Zurich.

Frey B.S, (1997), *Not Just for the Money: An Economic Theory of Personal Motivation*. Cheltenham, UK, Elgar.

Frey B. - Oberholzer-Gee F., (1997), "The Cost of Price Incentives: An Empirical Analysis of Motivation Crowding-Out", *American Economic Review*, 87:746-755.

Nelson R. W., (2002), "Equity or Intention: it is the Thought that Counts", *Journal of Economic Behavior and Organization*, 48: 423-430.

Pelligra V., (2002a), Fiducia R(el)azionale in P.L. Sacco and S.Zamagni (Eds.), *Complessità Relazionale: Fondamenti del comportamento economico*. Bologna: Il Mulino.

Pelligra V., (2002b), "Rispondenza Fiduciaria: Principi e Implicazioni per la Progettazione Istituzionale", *Stato e Mercato*, 65:330-353.

Pelligra V., (2002c), "Trust responsiveness: the self-fulfilling quality of trust", mimeo, University of East Anglia.

Pelligra V., (2003), "Consequences vs. Procedures: an experimental investigation", mimeo, Università di Cagliari.

Predergast C., (1999), "The provision of Incentives in Firms", *Journal of Economic Literature*, 37: 7-63.

Rabin M., (1993), "Incorporating Fairness in Game Theory", *American Economic Review*, 83: 1281-301.

Rotter J., (1966), "Generalized Expectancy for Internal versus External Control Reinforcement", *Psychological Monographs*, 80, 609.

Smith A., (1759/1976), *The Theory of Moral Sentiments*, Indianapolis, Liberty Classics.

Titmuss, R. (1970), *The Gift Relationship*, London: Allen and Unwin.

Williamson O., (1985), *The Economic Institutions of Capitalism*. New York. The Free Press.

* Texto presentado en la 44ª Reunión Anual de la Società Italiana degli Economisti, Salerno, 4-5 Octubre, 2003.

Traducción: Cecilia Kennedy.