

Capítulo Tres

PREFERENCIAS Y COMPORTAMIENTO

Los escritores políticos han establecido como máxima que al elaborar cualquier sistema de gobierno...debe suponerse que todo hombre es un *bellaco* y no tiene otro fin en todos sus actos, que su interés personal. Por este interés hemos de gobernarlo, y con él como instrumento obligatorio, a pesar de su insaciable avaricia y ambición, ha de contribuir al bien público.

-David Hume, *Ensayos: Moral y Política* (1742)

Volvamos al estado de naturaleza y consideremos a los hombres como si...acabaran de brotar de la tierra, y, de repente, (al igual que los champiñones) llegaran a la total madurez, sin ningún vínculo entre ellos.

-Thomas Hobbes, *El Ciudadano* (1651)

CULTIVAR MAÍZ ES UN GRAN negocio en Illinois. Usando tecnologías intensivas en capital y siguiendo planes de negocios diseñados por un ordenador, algunos granjeros cultivan mil acres o más, la mayoría en terrenos arrendados a varios propietarios. A mediados de la década de 1990 más de la mitad de los contratos entre los granjeros y los propietarios eran acuerdos de aparcería, y más de cuatro quintos de esos contratos estipulaban entre las partes una división de cincuenta – cincuenta de los cultivos. En el sur del estado, donde el suelo es en promedio menos fértil, hay condados en los que es común que los contratos otorguen al arrendatario dos tercios de la cosecha. En éstos condados hay pocos contratos de cincuenta – cincuenta o cualquier otra división distinta a los dos tercios de la cosecha entre las dos partes, a pesar de las variables consideraciones de la calidad de la tierra entre esos condados.

El cultivo de arroz en Bengala Occidental de mediados de la década de 1970 parece encontrarse a años luz del de Illinois. Los granjeros pobres y analfabetos de los pueblos aislados con carreteras que son intransitables la mayor parte del año y careciendo de comunicación electrónica, apenas vivían satisfaciendo las necesidades mínimas en terrenos que promedian los dos acres. Ya hemos visto (en el Prólogo) que estos comparten, sin embargo, una similitud con los granjeros de Illinois: la división entre los cultivadores y los propietarios era sobre un reparto del cincuenta – cincuenta en más de dos tercios de los contratos. (Ibn Batuta, cuya visita a Bengala también he mencionado en el prólogo, ha notado – y deplorado – exactamente la misma división de los cultivos seis siglos antes). Se observaron otros tipos de contratos, pero ninguno de ellos constituyó más del 8 por ciento del total.¹ Un ejemplo todavía más impresionante procede del Sur de los Estados Unidos después de la Guerra Civil, donde los contratos de aparcería dividían la cosecha equitativamente entre el dueño de la tierra y el tenedor, sin importar la calidad de la tierra o si el tenedor era un blanco libre desde el nacimiento o un esclavo recién liberado: “Esta forma de tenencia se estableció en cualquier parte del sur. Floreció con todo tipo posible de combinaciones de calidad del suelo y condiciones laborales” (Ransom y Sutch 1977:91, 215).

Lo que confunde de la aparcería de cincuenta – cincuenta es lo siguiente: la división equitativa de las cosechas significa que los tenedores de tierras fértiles tendrán pagos mayores por sus esfuerzos y otras entradas que los tenedores de tierras pobres. Pero si los tenedores están dispuestos a trabajar por menores ingresos en las

¹ Young y Burke (2001), Burke y Young (2000) y Bardhan (1984).

tierras menos buenas, ¿por qué los propietarios de las tierras buenas conceden la mitad de la cosecha a sus tenedores? La teoría económica convencional de aparcería predice que el dueño captará los retornos a la calidad de la tierra a través de variaciones en la división de la cosecha. (Stiglitz 1974). Pero Burke y Young (2000) muestran que los contratos de aparcería de Illinois permiten a los tenedores de tierra buena captar un tercio del retorno diferencial atribuible a la calidad de la tierra, transfiriendo efectivamente millones de dólares de los dueños a los granjeros.

Una interpretación posible de éstos hechos es que los granjeros y los dueños alrededor del mundo han dado con el cincuenta – cincuenta como una división aparentemente justa, y que los intentos de los dueños por captar todos los retornos de las tierras de alta calidad a través del uso de divisiones variables serían frustrados por la retaliación de los tenedores de las tierras. Si fuera cierta, esta interpretación sugiere que una predisposición a la justicia, así también como el deseo de castigar a aquellos que violan las normas locales, podrían ser motivos tan poderosos como la maximización de las ganancias y la persecución del propio beneficio.

John Stuart Mill (1965 [1848]) notó el impresionante patrón global de división equitativa en aparcerías, así como el cumplimiento local con las divisiones alternativas en las que no se observa el cincuenta – cincuenta. ¿La explicación de Mill? “La costumbre del país es la regla universal” (149). La costumbre puede ser la causa aproximada, pero esta explicación lleva a preguntarse: ¿Por qué cincuenta – cincuenta en lugar de cincuenta y dos – cuarenta y ocho? ¿Por qué los Bengalíes y los estadounidenses llegaron al mismo número? Sabemos por el análisis del juego de la división en el Capítulo 1, que *cualquier* división exhaustiva de la cosecha es un equilibrio de Nash Pareto eficiente: ¿Por qué éste particularmente? Aún más complejo: ¿Por qué persiste cuando los beneficios parecen ser enormes al ofrecerse menores porcentajes en tierras de alta calidad? Y cuando las divisiones cambian, como vimos que aconteció en Bengala Occidental en las décadas del 1980 y 1990, ¿Por qué todos cambian a la vez, reflejando el patrón de homogeneidad local y equilibrio puntuado que encontramos en el capítulo 2?

Si motivos como la justicia y la retribución, o simplemente la adhesión a las convenciones, invalidan el interés material propio en el ambiente altamente competitivo de la agricultura de Illinois, sería sabio reconsiderar los presupuestos económicos del comportamiento, los que convencionalmente han tomado al interés

individual –resumido en el término *Homo economicus* – como su fundamento. La necesidad de una segunda mirada al *Homo economicus* es clara cuando se consideran problemas de distribución como las aparcerías y otras situaciones de negociación en las que parecen destacar preocupaciones de equidad. Pero el problema es mucho más general, y el modelo canónico de comportamiento parece fallar frecuentemente aún cuando los temas de equidad parecen estar ausentes.

Consideremos el siguiente caso (Gneezy y Rustichini 2000). En todas partes, los padres algunas veces llegan tarde a recoger a sus niños en las guarderías. En Haifa, en seis guarderías escogidas al azar, se impuso una multa por retraso (en un grupo de guarderías de control no se impuso ninguna multa). La expectativa era mejorar la puntualidad. Pero los padres respondieron a la multa con retrasos aún más grandes: la fracción que recogía tarde a sus niños pasó a más del doble. Aún más sorprendente fue el hecho de que después de dieciséis semanas la multa se revocó, el aumento del retraso persistió, y no se vislumbraba una tendencia de regreso al status quo ante. Durante las veinte semanas del experimento, no hubo cambios en el grado de retraso en las guarderías del grupo de control.

Los autores del estudio piensan que la multa fue una pista contextual, que proveyó, inintencionadamente, información sobre el comportamiento apropiado.

El efecto fue el de convertir la tardanza, de violación de una *obligación* que los padres tenían algún inconveniente en respetar, en una mercancía con un precio que muchos estaban dispuestos a pagar. Llamaron a su estudio “Una multa es un precio” y concluyeron que al imponer una multa se calificó la interacción como una situación de mercado, en la que los padres estaban más que dispuestos a comprar tardanza. Revocar la multa no restauró el marco inicial de puntualidad como obligación, simplemente bajó el precio de la tardanza a cero. El hecho de que los incentivos monetarios para la puntualidad indujeran a una tardanza mayor es tanto una oposición a las predicciones del modelo económico estándar como una cuestión de relevancia general al problema de diseñar contratos efectivos y políticas económicas. En términos de Hume, las guarderías de Haifa diseñaron una constitución para bellacos, y aparentemente produjeron bellacos en lugar de comportamientos mejorados.

Las debilidades del modelo convencional sugerido por la confusión de la división cincuenta – cincuenta de los cultivos y lo contraproducente de multar a los

padres de Haifa son evidentes en terrenos de interés económico más convencionales, como los mercados laborales, la producción en equipo, el acatamiento de los impuestos, la protección del ambiente común local, y otras formas de disposición de los bienes públicos. Allí se incluyen los motivos de equidad en el ajuste de los salarios y otros intercambios (Bewley 1995, Blinder y Choi 1990). Igual de desconcertante para el paradigma estándar es el hecho de que los individuos se preocupan por votar, dado que la probabilidad de que su voto sea decisivo es extremadamente pequeña. Cuando sí votan, el apoyo que usualmente le dan a la transferencia de ingresos para los pobres mediante impuestos, aún entre aquellos que son suficientemente ricos o con altas posibilidades de ascenso social como para beneficiarse directamente de ellos, resulta igual de desconcertante (Fong 2001, Gilens 1999). Finalmente, estudios en Continental Airlines, Aceros Nucor y otras compañías han hallado que los incentivos grupales son efectivos aún donde los beneficios se dividen entre un gran número de personas por lo que la ganancia de cada individuo según su propio esfuerzo es insignificante (Hansen 1997, Knez y Simester 2001).

Buscando una base de comportamiento más adecuada para la economía y otras ciencias sociales, en éste capítulo delinearé una investigación reciente con el fin de presentar una reformulación del método estándar, uno que retiene el papel central a las preferencias individuales y creencias, y obliga a explicar cómo actúa la gente, a la vez que corrige el modelo convencional de tres formas.

Primero, muchos comportamientos se explican mejor por lo que se ha llamado *preferencias sociales*: al elegir como actuar, los individuos comúnmente toman en cuenta no sólo las consecuencias de sus acciones para ellos mismos, sino también para otros. Más allá de eso, en general no sólo se preocupan por las consecuencias sino también por las intenciones de los otros actores. Un ejemplo importante de preferencias sociales son los motivos de *reciprocidad*, de acuerdo con los cuales la gente es generosa con las personas que se han portado bien (con ellos u otros), mientras que castigan a aquellos que no se han comportado bien. Los motivos de reciprocidad inducen a la gente a actuar de esta manera aún en situaciones (como las interacciones únicas) en las que los comportamientos de generosidad y castigo son personalmente costosos y no llevan la expectativa de recompensa subsiguiente o indirecta. (Estos casos son ejemplos de lo que yo llamo *reciprocidad fuerte*, para distinguir este comportamiento de la reciprocidad con expectativa de una recompensa futura, algunas veces llamada

altruismo recíproco). Otras preferencias sociales que deben considerarse son la *aversión a la inequidad*, *envidia* (o fastidio), y *altruismo*.

Contrario a ello, el supuesto convencional es que el comportamiento individual se explica completamente por lo que se llama imprecisamente interés por uno mismo que yo me refiero como *preferencias orientadas a sí mismo definidas por los resultados*. De acuerdo con éste punto de vista, nuestras preocupaciones no se extienden ni a los resultados experimentados por otros, ni a los procesos que generan resultados. F. Y. Edgeworth, uno de los fundadores del paradigma neoclásico, expresó su posición en su libro *Física Matemática* (Edgeworth 1881:104): “El primer principio de la economía es que cada agente actúa solamente en su propio interés” El interés por uno mismo no se presume por la racionalidad (uno puede tener preferencias transitivas y completamente altruistas o masoquistas), pero se trata comúnmente en economía como axiomática (y algunas veces se confunde con racionalidad). Así, mientras que el interés por uno mismo no está implicado formalmente en el método convencional, generalmente en la práctica se presume. La presunción adquiere un considerable poder de predicción en situaciones estratégicas cuando toman la forma de lo que llamo el *axioma del interés propio*, es decir, el interés personal individual conectado con la creencia que otros también están motivados por el interés por uno mismo.

Segundo, los individuos son *agentes adaptativos - seguidores de reglas*. Con esto quiero decir que economizamos en nuestros limitados recursos cognitivos actuando de acuerdo a las reglas de oro evolucionadas. El término “racionalidad acotada” es usado algunas veces para describir los límites cognitivos de actores humanos reales, pero yo no lo uso, porque sugiere irracionalidad. No es la limitación de nuestra racionalidad lo que quiero remarcar, sino nuestra limitada capacidad y predisposición para comprometernos en ejercicios cognitivos extremadamente costosos y complejos. Entre éstas reglas de comportamiento evolucionadas existen prescripciones éticas que gobiernan las acciones hacia los otros, es decir, *normas sociales*, cuya conformidad a las mismas es valorada por el actor (ej. la norma está internalizada) y apoyada por la sanción social. Este enfoque contrasta con la visión convencional en la cual el comportamiento es el resultado de procesos cognitivos individuales muy exigentes dirigido a temas tanto evaluativos como causales (¿es este estado deseable? ¿puedo lograrlo?). Este punto de vista convencional *centrada en la cognición individual* excluye el comportamiento basado en cosas como reacciones viscerales (como disgusto, miedo, o debilidad de voluntad), hábitos o reglas de oro

evolucionadas, y presupone (contra un importante cuerpo de evidencia) que los individuos están capacitados y predispuestos a hacer inferencias bastante avanzadas sobre lo que harán los otros y sobre la forma en que funciona el mundo.

Tercero, los comportamientos dependen del contexto en tres sentidos. Las pistas situacionales se usan para determinar los comportamientos apropiados en cualquier escenario dado. Más allá de eso, evaluamos los resultados desde un punto de vista particular, es decir nuestro estado actual o el estado experimentado por un miembro de nuestro grupo de referencia. Finalmente, las instituciones sociales influyen en quién conocemos, para hacer qué, y con qué recompensas. Como resultado de lo anterior, nuestras motivaciones se forman a través del proceso de transmisión cultural o genética, que fue presentado en el capítulo 2. Así, nuestras *preferencias son situacionalmente específicas y endógenas*. Si nuestras experiencias resultan en cambios durables de nuestras preferencias, se dice que estas son endógenas, lo que sucederá si las experiencias afectan o al aprendizaje social (a través de un tiempo prolongado), o a la herencia genética. Esto puede compararse con preferencias situación - dependientes o estado -dependientes que son invariables en el tiempo (a través del tiempo uno se comporta de la misma manera en la misma situación). Dado que las preferencias endógenas implican cambios de aprendizaje o cambios genéticos, el comportamiento ante situaciones iguales, cambia a través del tiempo.

Este enfoque contrasta con la visión convencional que las preferencias no dependen del estado actual de uno y son invariables o cambian solamente por influencias exógenas para el problema bajo investigación. George Stigler y Gary Becker (1977) explicaron este punto de vista en su *ensayo De Gustibus Non Est Disputandum*: “Uno no discute sobre los gustos por la misma razón que uno no discute acerca de las Montañas Rocosas – ambos están ahí y estarán también el año próximo, y son lo mismo para todos los hombres”(76). Ellos repetían, en términos menos poéticos, el punto de Hobbes acerca de los champiñones.

Nadie toma, por supuesto, los supuestos convencionales literalmente. Edgeworth observó que el supuesto del interés en sí mismo es literalmente verdadero sólo en situaciones límite (“contrato y guerra”), y Hume, en la oración que sigue inmediatamente al primer epígrafe de éste capítulo, reflexionó que “es extraño que una máxima deba ser verdadera en la política, que de hecho es falsa”. Hobbes invocó una analogía deliberadamente caprichosa para extraer la formación social de las

preferencias como parte de un experimento de pensamiento, no como una descripción de personas reales.

Mientras se reconoce que los supuestos estándar son a menudo violados empíricamente, la mayoría de los economistas han compartido el apoyo de Becker y Stigler al modelo canónico simple de preferencias exógenas e interés por uno mismo. La amplia aceptación de sus principios – no como verdades empíricas sino como aproximaciones lo suficientemente cercanas para ser atajos analíticos útiles –, se explica en parte por la contribución sustancial para la disciplina intelectual y la claridad. Los supuestos estándar proveen un marco intelectual común resistente a explicaciones ad hoc sobre la base de diferencias individuales no observadas empíricamente o cambios en los gustos a través del tiempo. Abandonar el modelo estándar abre la puerta a explicaciones o comportamientos sobre la base de conceptos vagos como “ingreso psíquico” o “espíritus animales”.

Para que una base de comportamiento sea una contribución a la ciencia social en lugar de una invitación a explicaciones ad hoc, necesitamos más información empírica acerca de las preferencias y cómo llegamos a ellas, así como modelos de comportamiento más adecuados bajo supuestos de preferencia menos restrictivas. La extraordinaria producción de hallazgos empíricos por parte de economistas experimentales y del comportamiento y otros científicos sociales en años recientes, han hecho dicha reformulación posible y atrasada. Aquí y en capítulos siguientes hago uso extensivo de los resultados experimentales. La razón es que éste método, relativamente nuevo en economía, ha permitido por primera vez la prueba de hipótesis bien formuladas relacionadas con asunciones de la economía en escenarios controlados.

En la próxima sección presento lo que llamo una interpretación del comportamiento de preferencias y acciones racionales, seguido por una revisión de un número de anomalías empíricas en el tratamiento convencional de las preferencias. Luego paso a investigaciones recientes sobre preferencias sociales, mostrando resultados experimentales y dos nuevas funciones de utilidad. Pospongo hasta los capítulos 7, 11 y 13 el diseño formal de cómo evolucionan las preferencias, por qué la gente a menudo se adhiere a normas éticas, y por qué son comunes otros motivos como la generosidad y la justicia.

PREFERENCIAS, RAZONES Y COMPORTAMIENTOS

Cuando los individuos actúan, generalmente intentan *hacer* algo, sea inteligente o no. Una implicación es que los propósitos de los individuos y su comprensión acerca de cómo deben llevarlos a cabo, junto a las limitaciones y los incentivos planteados por las reglas sociales y las capacidades individuales, son ingredientes claves cuando se explican las acciones individuales. Lo que hace la gente depende en cualquier situación de sus preferencias individuales y de sus creencias.

Las *creencias* son las comprensiones del individuo sobre de la relación existente entre una acción y un resultado. En muchos casos las creencias entran de forma trivial en la elección de situaciones por lo que no se abordan explícitamente: por ejemplo, generalmente asumimos, en juegos simples, que la gente conoce los pagos que resultan de sus acciones. En otras situaciones – particularmente en interacciones estratégicas sin estrategias dominantes –, las creencias pueden volverse muy importantes: el efecto que yo asista a una reunión puede depender de quién más este allí entonces mi decisión de asistir o no dependerá de mis expectativas acerca de quién más asistirá, lo que también dependerá de *sus* creencias acerca de los otros que asistirán, y así sucesivamente. En otras situaciones la estructura de la interacción puede ser ambigua y entendida de forma diferente por distintos jugadores. En estas situaciones tiene una importancia central cómo llegamos a creer lo que creemos y cómo actualizamos nuestras creencias a la luz de nuestra experiencia.

Las *preferencias* son las razones del comportamiento, es decir atributos de los individuos – diferentes de las creencias y las capacidades-, que dan cuenta de las acciones que se toman en una situación dada.² Así, las preferencias incluyen una mezcla heterogénea: gustos (gusto o disgusto por comidas), hábitos, emociones (como vergüenza o ira) y otras reacciones viscerales (como el miedo), la forma en que los individuos interpretan situaciones (o más limitado, la manera en que elaboran una decisión), compromisos (como promesas), normas impuestas socialmente, propensiones psicológicas (por agresión, extroversión, etc), y una relación afectiva con los otros. Decir que las personas actúan según sus preferencias significa sólo que el conocimiento de las preferencias ayudará a explicar las acciones (aunque no necesariamente la explicación que sería dada por el actor, ya que es bien conocido

² Un término más preciso para esta concepción de las preferencias puede ser la expresión difícil de manejar sugerida por Novell – Smith (1954): “Actitudes pro y contra”.

que los individuos algunas veces no son capaces o no están dispuestos a dar una explicación de ese tipo.)³

Esta interpretación de las preferencias como “razón para el comportamiento” puede contrastarse con dos enfoques convencionales. El primero establece que los individuos buscan maximizar sus utilidades, equiparando la utilidad al bienestar, el placer, la felicidad, en la tradición de Jeremy Bentham y los utilitaristas del siglo diecinueve. Contrario a ello, para el enfoque más reciente de las preferencias, un orden de preferencias no es más que una *descripción* completa de un comportamiento consistente y cualquier relación con un cálculo hedonístico es gratuita. Ninguno de los planteamientos es completamente adecuado.

Si nuestro objetivo es explicar el comportamiento, el método de la preferencia revelada es vacío porque no dice nada sobre los motivos y las razones: mientras que rara vez son suficientes para una explicación es difícil que no contengan información. El punto de vista de la preferencia revelada alguna vez atrajo adeptos impresionados por el hoy anticuado orden metodológico acerca de la imposibilidad de conocer los estados subjetivos, dado que un método científico debía basarse en comportamientos observables. En contraste el método utilitarista es sustantivo; los estados subjetivos, esenciales para este punto de vista – placer, dolor, satisfacción, ansiedad, y otras experiencias hedónicas-, son ahora un campo activo de estudio y de medición científica. Pero, tratar al comportamiento como sinónimo de la búsqueda del bienestar es engañoso: las razones para nuestro comportamiento incluyen adicciones, debilidad de la voluntad, miopía, y otros aspectos disfuncionales bien documentados del comportamiento humano. El hecho que el mismo término – utilidad – se use convencionalmente en ambos casos, como una explicación del comportamiento y como un estándar para evaluar resultados sociales, ha forzado a los economistas a tomar un punto de vista indebidamente limitado del comportamiento y de la evaluación social.

Para revisar esto, junto al conjunto de acciones viables y de sus resultados asociados, las creencias y las preferencias dan una explicación de la acción individual. Recordemos que he definido a las instituciones como las leyes a nivel de población, normas informales, y convenciones que otorgan una estructura durable a las

³ Ver Nisbert y Wilson (1977). Shafir, Simonson, y Tversky (2000) dan una interpretación de lo que llaman “Elección basada en la razón”, similar a la de Novell – Smith y aquí.

interacciones sociales. En términos teóricos de juegos, una institución es un juego (el cual como hemos visto en el capítulo 1, puede también ser el resultado de un juego subyacente), las preferencias son la evaluación de los pagos y las creencias son el entendimiento que tienen los jugadores de las consecuencias que tiene su estrategia en términos de pagos esperados (ej. su comprensión del juego y de la estructura de pagos además de las probabilidades de las acciones de los otros).

Como las preferencias, las creencias, y las instituciones se confunden fácilmente. Consideremos un caso concreto. En muchos países, la práctica común de conducir del lado derecho de la carretera es una institución; es una convención, es decir, un equilibrio del Juego del Aseguramiento, y la convención está apoyada por leyes. En estos países es una mejor respuesta conducir por la derecha y también es ilegal hacerlo de otra forma. Las personas no *prefieren* conducir por la derecha per se, prefieren evitar choques y multas y donde todos los demás conducen por la izquierda sin violar la ley, ellos conducirán también por la izquierda. La creencia que otros conducirán por la derecha sustenta la institución de conducir por la derecha, lo que a su turno sustenta la creencia. Las creencias y las preferencias son hechos de los individuos que sustentan este equilibrio particular, mientras las instituciones – representadas en este caso por el equilibrio de conducir por la derecha – son hechos de grupos de personas.

Una versión del marco de las creencias y las preferencias que llamare “convencional” ha provisto de la base del comportamiento para la economía y se aplica cada vez más en todas las ciencias sociales. El comportamiento de un individuo se representa utilizando una función de utilidad: $U: U(x, y, z)$. Los argumentos para U – $x, y, y z$ – describen un *estado* que puede ser una simple lista de productos consumidos o formulaciones más complejas como una cerveza fría en una tarde caliente dentro de tres días en la compañía de amigos en una sociedad islámica que prohíbe el consumo de alcohol. La función de utilidad se toma como una representación numérica tal que los valores mayores de U son elegidos sobre (digamos son preferidos) los valores menores, el estado (x, y, z) es elegido sobre (x', y, z) si $U(x, y, z) > U(x', y, z)$.

La función de utilidad es *completa* lo que significa que cada estado puede ordenarse por una relación de preferencia o de indiferencia respecto a otro estado. El orden es también *transitivo* lo que significa que la ordenación que produce no incluye

ordenaciones inconsistentes como (x, y, z) a preferencia de (x', y, z) , lo que es preferido a (x'', y, z) , pero (x'', y, z) se prefiere a (x, y, z) . Finalmente, la función de utilidad se presume (por lo general implícitamente) como *invariable en el tiempo* dentro del periodo relevante: cuando, digamos, los precios cambian exógenamente el individuo responde a los nuevos precios y no sólo a los cambios coincidentes en la función de utilidad. Cuando los individuos actúan de acuerdo a una función de utilidad completa y transitiva se dice que son *racionales*.⁴ Otras formas de actuar – inconsistencia de elección inducida por capricho o preferencias incompletas sobre resultados inimaginablemente horribles, por ejemplo – no se consideran irracionales. Son simplemente formas de acción que no están cubiertas por este modelo y tal vez sería mejor considerarlo como *no racional*.

El modelo convencional se extiende usualmente para cubrir el riesgo y la incertidumbre. Se dice que existe *riesgo* si una consecuencia de una acción en el conjunto de elecciones del individuo es un conjunto de resultados posibles, cada uno ocurriendo con una probabilidad *conocida*. Al contrario, si una o más acciones abiertas al individuo pueden causar más de un resultado y las probabilidades de los mismos son *desconocidas*, existe *incertidumbre*. Ambos son aspectos ubicuos de elección. Decidir si alquilar una casa en la playa sabiendo que con la probabilidad p lloverá es un ejemplo de riesgo. En estos casos el individuo se presume que quiere maximizar la *utilidad esperada*. La utilidad esperada de una acción es la utilidad asociada con cada una de las posibles consecuencias de la acción multiplicada por la posibilidad de que ocurra: $U(\text{casa en la playa}) = (1 - p)U(\text{casa en la playa con sol}) + pU(\text{casa en la playa con lluvia})$.

La maximización de la utilidad esperada requiere más que el simple ordenamiento de cada estado posible (eso es suficiente para determinar el comportamiento bajo incertidumbre), ya que utiliza información acerca de cuanto mejor es un estado que otro. En un trabajo pionero en la teoría de juegos, John von Neumann y Oskar Morgenstern (Neumann y Morgenstern 1944) demostraron que una utilidad esperada que maximiza las opciones del individuo es invariable ante una transformación aditiva o lineal de la función de utilidad. (Esto significa que si el comportamiento del individuo se describe por la función de utilidad u , entonces su comportamiento se describe por cualquier función de la forma $v = \alpha + \beta u$ donde $\beta >$

⁴ A veces se imponen otras restricciones racionales. Por ejemplo, el débil axioma de la preferencia revelada requiere que si (x, y, z) es preferido a (x', y, z) , entonces (x, y, z, a) será preferido a (x', y, z, a) .

0.) Lo que se llama *utilidades von Neumann – Morgenstern* contienen esta restricción. Estas ya han hecho dos apariciones sin aviso en el capítulo 1: en el tratamiento del riesgo dominante y cuando normalicé los pagos asociados con las posiciones en retirada en juegos de conflicto de interés. Las utilidades von Neumann – Morgenstern presentan cardinalidad de los estados de *un individuo particular* pero no *entre* individuos; indican cuánto mejor es la playa con sol comparada a la playa con lluvia *para usted*, pero no cuánto mejor es *para usted* que *para mi*. Todos los pagos que se usen de aquí en adelante serán *utilidades von Neumann-Morgenstern* a menos que se especifique lo contrario.

En el caso de la incertidumbre, el peso de la probabilidad conocida se reemplaza por estimaciones subjetivas del individuo acerca de las probabilidades desconocidas. Generalmente, se presume que los individuos modifican sus estimaciones con base en su experiencia reciente mediante un proceso llamado *actualización Bayesiana*; el reverendo Thomas Bayes (1702 – 1761) fue de los primeros escritores de la teoría de la probabilidad. El enfoque Bayesiano de la acción racional presume que la toma de decisiones del individuo bajo incertidumbre se basa en la maximización de la utilidad esperada basada, a su vez, en las probabilidades subjetivas actualizadas de esta manera. (el enfoque Bayesiano obviamente presupone las utilidades *von Neumann-Morgenstern*). En la práctica, la diferencia entre riesgo e incertidumbre es a menudo borrosa, excepto en casos límite donde las probabilidades verdaderamente conocidas están involucradas como mecanismos de distribución que son escogidos al azar por el lanzamiento de una moneda.

Una aplicación importante de estas ideas es el concepto de *aversión al riesgo* medido por el grado de concavidad de una función de utilidad $U(W)$, donde W es la riqueza del individuo. La intuición es que si la utilidad marginal de la riqueza está declinando rápidamente con la riqueza, como será el caso para una función de utilidad “muy cóncava” entonces uno valorará \$75,000 con certeza como mucho más que una probabilidad aún mejor de \$50,000 ó \$100,000. Así, un individuo cuya utilidad es cóncava en riqueza puede tener aversión a la lotería entre dos premios si puede tener, en su lugar, cierto premio que iguala el valor esperado de la lotería. Por esta razón, una medida del grado de aversión al riesgo es $-U''/U'$, llamada medida

Arrow-Pratt.⁵ Un individuo es *neutral al riesgo* si la utilidad es lineal en riqueza o $U'' = 0$; $U'' > 0$ implica que es un *amante del riesgo*.

Una segunda extensión esencial es la de elecciones sobre estados en diferentes fechas. Esto se logra descontando los estados futuros con el factor δ de descuento, que es una medida inversa del grado al que descontamos eventos futuros debidos a la miopía, la poca probabilidad de sobrevivir a alguna fecha futura y otras razones.⁶ Para una persona que valora los estados futuros tanto como los estados actuales, $\delta = 1$ mientras que para individuos más orientados al presente, $\delta < 1$. De acuerdo con el enfoque del *descuento de utilidad*, δ se define tal que un individuo sea indiferente entre sumar x a su consumo y en el tiempo t y sumar otros incrementos x' , en n periodos más tarde, en $t + n$ si

$$U(y + x) \delta^t + U(y) \delta^{t+n} = U(y) \delta^t + U(y + x') \delta^{t+n} \quad (3.1)$$

Así, extendido para cubrir el riesgo y la elección inter-temporal, el modelo convencional capta el importante aspecto intencional del comportamiento humano y combina una amplia aplicabilidad con la posibilidad de hacerle un seguimiento formal. A primera vista parece imponer pocas restricciones sustantivas al análisis del comportamiento, más allá de la exclusión de los casos tal vez poco importantes de lo incompleto e inconsistente recién mencionados. Pero esto no es correcto: la formulación de arriba es una teoría sustantiva del comportamiento y contiene fuertes demandas sobre el tipo de cosas que toma en cuenta la gente y cómo lo hace. Este modelo no sale bien librado a la luz de investigaciones empíricas recientes sobre el comportamiento.

PREFERENCIAS SITUACIÓN DEPENDIENTES

Una de los errores mejor documentados del modelo convencional surge por las preferencias (y por lo tanto los comportamientos) que son *situación - dependientes* en el siguiente sentido. Supongamos que ω_i es un vector que representa el estado i (ej. uno descrito como el (x, y, z) de arriba), un elemento del conjunto de estados posibles Ω , y $U_i(\omega_i)$ es la utilidad asociada con el estado $\omega_i \in \Omega$ para un estado actual de la

⁵ Ver Mas –Colell, Whinston, y Green (1995) para mayor elaboración.

⁶ El factor de descuento $\delta = 1/(1 + r)$ donde r es la tasa de preferencia temporal.

experiencia individual ω_i . Sea $U_i(\omega)$ que representa la clasificación de las preferencias individuales de todos los estados posibles cuando ese individuo está en estado i . Ahí las preferencias son situación - dependientes si la clasificación del mismo individuo en diferente estado, dado por $U_k(\omega)$ difieren de aquellas dadas por $U_i(\omega)$ para algún i y k . La dependencia situacional también es llamada dependencia de estado, pero yo uso la primera en reconocimiento a la literatura sustancial de la psicología sobre la importancia de las situaciones como influencia para el comportamiento.

Un ejemplo importante de dependencia situacional, llamado *aversión a la pérdida*, surge porque las personas valoran (negativamente) las pérdidas más que las ganancias equivalentes. El tamaño del coeficiente de la aversión a la pérdida es sorprendentemente grande: estimaciones de ambos experimentos y en ambientes naturales encontraron que la inutilidad de una pérdida pequeña es entre dos y dos y media veces la utilidad de una ganancia pequeña. La función de utilidad está marcadamente ligada al status quo (y el nexo se mueve cuando cambia el status quo). A esto está fuertemente asociado el efecto dotación: el precio mínimo que inducirá a un individuo a vender algo que ahora posee es sustancialmente mayor que el precio máximo que estará dispuesto a pagar para adquirir el mismo bien. (La aversión a la pérdida y el efecto dotación son ejemplos de una clase más amplia de efectos situación - dependientes, llamados sesgo por el *status quo*).

La aversión a la pérdida y el efecto dotación han sido documentados extensamente en experimentos hechos por economistas y psicólogos, y dan explicaciones plausibles de anomalías importantes en la economía cotidiana. Por ejemplo, el hecho que los retornos de las acciones en Estados Unidos han excedido consistentemente los retornos de los bonos por un amplio margen es un interrogante sobresaliente en economía. Una vez se pensó que era un resultado de la aversión al riesgo entre inversionistas, pero un simple cálculo (Mehra y Prescott 1988) mostró que el nivel necesario de aversión al riesgo para explicar la diferencia es inverosímilmente grande. Para que la aversión al riesgo se tome en cuenta en el problema del retorno de las acciones, los inversionistas deberán ser indiferentes entre una mejor probabilidad de \$50,000 y \$100,000 y una cosa segura de \$51,209. Una versión más convincente (Benartzi y Thaler 1995) sostiene que los inversionistas no tienen aversión a las variaciones de los retornos por sí mismos (después de todo la mayoría son bastante ricos), sino que reaccionan fuertemente ante la perspectiva de

pérdidas y los retornos de las acciones en un año son negativos con mayor frecuencia que los retornos de los bonos.

Los interrogantes a la interpretación de aversión a las pérdidas en los retornos de las acciones pone en claro que una formulación precisa de la aversión a la pérdida y otros aspectos de la situación dependencia requieren de un tratamiento explícito de las dimensiones de tiempo; si los inversionistas tienen un horizonte de tiempo de cinco años experimentarían pocos retornos negativos, por lo que la explicación de la aversión a las pérdidas implica un horizonte de tiempo particular, evidentemente de preferencia uno corto. Un individuo que experimenta una pérdida eventualmente tratará a la nueva situación como un status quo. Sabemos que las personas que anticipan que una discapacidad física severa será insoportable, a menudo estarán bastante satisfechas con la vida después de vivir con la discapacidad por algunos años. Un determinante situacional de las preferencias bien documentado es la simple exposición (Zajonc 1968). Las personas comienzan a valorar más las cosas (por ejemplo la comida) si han estado expuestas a ellas por más tiempo. Las ratas no son diferentes: las que han crecido con Mozart prefieren su música a la de Schoenberg (Cross, Halcomb, y Matter 1967). Algunas veces las preferencias se adaptan a las situaciones casi en forma instantánea – ¡los estudiantes en experimentos de efecto dotación se relacionan con las tazas de café que reciben en cuestión de minutos! – pero la demora en la adaptación es considerablemente mayor en muchos casos.

La situación – dependencia, en la forma de aversión a las pérdidas, efectos dotación, y endogeneidad de las preferencias a largo plazo, bajo ningún concepto agota los defectos empíricos del modelo convencional. Como el supuesto de la independencia de la situación, el tratamiento convencional de la elección intertemporal es impresionantemente contra intuitivo y contradicho fuertemente por evidencia sobre el comportamiento.⁷ Imaginemos que a usted le es indiferente una comida en su restaurante favorito ahora y dos de esas comidas dentro de un año. De acuerdo a la ecuación (3.1) también le será indiferente una comida (llamada x) dentro de veinte años y dos comidas (eso es x') dentro de veintiún años. Para ver esto, nótese que esta relación de indiferencia puede expresarse de forma equivalente (divida ambos lados de (3.1) por δ') como

⁷ Este párrafo se extrae de Loewenstein y Prelec (2000).

$$U(y + x) - U(y) = \{U(y + x') - U(y)\} \delta^n$$

Así, la diferencia en la utilidad producida por el retraso de las dos comidas no depende de cuánto tiempo real pasa sino sólo de la cantidad de tiempo transcurrido entre la primera (una comida) y la segunda (dos comidas). Esto que se llama *propiedad de estacionalidad* del modelo de utilidad descontada es una analogía temporal a la independencia de estado: *cómo uno evalúa los estados se asume que no depende desde donde los está evaluando*. Esto no sólo es contra-intuitivo, lo contradice extensa evidencia experimental y otra (de manera interesante, para otros animales como para humanos). Para la mayoría de las personas, como lo sugiere el ejemplo, el retraso de un año sobresale más si ocurre más temprano que tarde sugiriendo lo que se llama una *función de descuento hiperbólico*, según la cual un estado en el año t no se descuenta a la tasa δ^t sino a la tasa

$$\delta(t) = (1 + \alpha t) \delta^{-\alpha t} \quad \text{con } \alpha, \beta > 0 \quad (3.2)$$

lo que para valores grandes de α indica que el valor de los estados futuros desciende rápidamente en el futuro cercano, por tanto el descenso es marcadamente atenuado (por ejemplo, usted puede ser bastante impaciente en esperar un año para su comida favorita pero menos impaciente para evaluar las consecuencias a largo plazo del calentamiento global).⁸ Los agentes que usan los descuentos hiperbólicos mostrarán un comportamiento de preferencias invertido: de dos premios A y B de cantidades diferentes y con ocurrencia en fechas futuras distintas, uno preferirá A sobre B en el presente pero, con el paso del tiempo preferirá B sobre A. Un agente con descuento hiperbólico podrá tomar una comida ahora sobre dos comidas dentro de un año, pero también elige las dos comidas dentro de veintiún años sobre una comida dentro de veinte. Pero si este es el caso, tras diecinueve años, ese agente con descuento hiperbólico elegirá una comida antes que las dos comidas más tarde, es decir, invertirá su elección. Un número de estudios (examinados en Angeletos, Laibson, Repetto, Tobacman, y Weinberg 2001) sugieren que el enfoque del descuento hiperbólico provee mejores predicciones que el enfoque convencional de comportamiento individual de ahorro, responsable de las observaciones empíricas sobre los incrementos en el consumo por incrementos predecibles en los ingresos y la marcada reducción del consumo luego de pensionarse.

⁸ El distanciamiento del descuento constante está regido por α ; puede confirmar que como α va a cero en la ecuación (3.2) reproduce la función de descuento exponencial estándar $\delta(t) = e^{-\beta t}$

Como en el caso de la elección intertemporal, regularidades empíricas bien establecidas son anómalas desde el punto de vista del análisis convencional de opciones de la utilidad esperada en presencia de riesgo. Recordemos que este marco requiere que los individuos evalúen las acciones que podrán tomar de acuerdo a la suma lineal de la probabilidad de cada posible consecuencia que ocurra, multiplicada por las utilidades asociadas con cada consecuencia. Así, eventos que ocurren con probabilidades arbitrariamente pequeñas deben ser tratados virtualmente sin diferencia de los eventos que ciertamente no ocurrirán. Pero sabemos que las personas no evalúan de esta forma las loterías sobre eventos con riesgo: un evento que ocurrirá con certeza es visto de modo bastante diferente que otro que pasará con probabilidad $(1 - \epsilon)$ sin importar como de pequeño sea ϵ . A la inversa, saber que uno no es VIH positivo es difícilmente lo mismo que saber que uno puede ser VIH positivo pero con una probabilidad arbitrariamente pequeña de ϵ . Paul Samuelson (1963) llamó a este suceso el problema “*epsilon no es cero*”.

Aparece un segundo problema: si la aversión al riesgo (medida por la concavidad de la función de utilidad en la riqueza) se utiliza para explicar porqué las personas retiran apuestas cuando están en juego cantidades en el rango entre \$0 y \$1,000, no puede explicar porqué *cualquier* apuesta es virtualmente aceptada cuando están en juego cantidades grandes. Un economista que ha observado a un individuo rechazar la oportunidad de lanzar una moneda con la posibilidad de ganar \$1010 o de perder \$1000 invocará la aversión al riesgo como explicación. Pero Matthew Rabin (2001) señaló que el nivel necesario de aversión al riesgo para explicar esta elección implicará también que el mismo individuo rechazará lanzar la moneda por una pérdida de \$80,000 o una ganancia de \$349,400. El problema es que para cantidades pequeñas en juego una función de utilidad cóncava es aproximadamente lineal, y la cantidad de concavidad necesaria para explicar porqué pequeñas sumas en juego son algunas veces rechazadas implica que la mayoría de las apuestas sobre grandes suma en juego – aún algunas muy lucrativas en términos de valor esperado – nunca serán aceptadas.

La idea de que reducir marcadamente la utilidad marginal de la riqueza que surge de la función de utilidad cóncava hará que un individuo decline tomar riesgo sobre grandes cantidades en juego es seguramente correcta. Pero los dos problemas de arriba sugieren que la concavidad por sí sola no puede explicar el comportamiento frente al riesgo. El primero es familiar: el enfoque convencional se abstrae de la

aversión a la pérdida. La segunda es más profunda: aún si la función de utilidad fuera continuamente diferenciable (no ligada al estado de status quo como sería el caso si la aversión a la pérdida estuviera presente), su concavidad fracasa en captar las razones que tienen las personas para desear evitar el riesgo y las emociones que experimentan de cara al riesgo. Entre ellas están la ansiedad y el miedo cuando no saben qué pasará o la posibilidad de lamentarlo (o la vergüenza) de haber tomado una oportunidad que ex post no valió la pena. El modelo fracasa al explicar las razones por las que las personas con una riqueza muy limitada participan en actividades con mucho riesgo como los juegos de apuestas: no es probable que sus funciones de utilidad sean *convexas* en riqueza, y, si lo son, entonces surge la pregunta de porqué los mismos individuos también compran seguros. Una explicación más plausible sobre el juego de apuestas y también sobre conducir a grandes velocidades, es que algunas personas disfrutan al tomar algunos *tipos* particulares de riesgos.

Las utilidades situación -dependientes, así como los resultados específicos del método de maximización de utilidad esperada con respecto al riesgo y el método de utilidad descontada con relación a la elección intertemporal, sugieren que se necesita un punto de vista con mayor fundamento empírico sobre de las razones del comportamiento. Daniel Kahneman, Amos Tversky, Richard Thaler, y sus coautores han sugerido una serie de reformulaciones llamada *teoría de la prospección* (los documentos claves se presentan en Kahneman y Tversky 2000). Su mayor contribución es tomar en cuenta cuatro aspectos de la elección que no están bien tratados en el paradigma convencional. El primero es el problema (mencionado arriba) de que las personas no evalúan las decisiones con riesgo de acuerdo con la hipótesis de la utilidad esperada: ellas sopesan la importancia de los eventos improbables. La segunda es tomar en cuenta el *marco*, es decir, el hecho de que los resultados equivalentes son tratados en forma diferente dependiendo de la manera en la que se describen tanto los resultados como el entorno de la decisión. Una de las razones del comportamiento situación – dependiente es que las situaciones a menudo enmarcan opciones de una manera particular. (Ejemplos de ello serán dados en la siguiente sección.) Tercero, Kahneman y otros, volviendo a un aspecto del utilitarismo clásico, reintrodujeron medidas sustantivas como la utilidad hedónica realmente experimentada.

Cuarto, la teoría prospectiva desarrolló un marco conceptual para tratar los comportamientos situación – dependientes. Esta reformulación fundamental dice

que si la función de utilidad se utiliza para explicar el comportamiento actual, sus argumentos deberían ser los *cambios de estado* o de *eventos* en vez de estados. Así, el valor que los individuos le dan a los estados depende de la relación de los estado con el status quo (o posiblemente algún otro estado de referencia, como un nivel de aspiración o los estados que disfrutan sus semejantes). Estudios experimentales y empíricos sugieren que la llamada función valor resultante tiene tres características que se ilustran en la figura 3.1, es decir, ese valor es definido en los cambios en la riqueza y no en los niveles, que la función de valoración está ligada al status quo con un coeficiente de aversión a la pérdida cercano a dos o un poco más (la función inmediatamente a la izquierda del status quo es dos veces más empinada que la de la derecha) y que el valor marginal de los cambios disminuye cuando aumenta la distancia al estatus quo, presentando efectos marginales aún menores en la valoración que hace el individuo del evento para que la función de valor sea convexa en las pérdidas y cóncava en las ganancias (el llamado *efecto reflejo*).

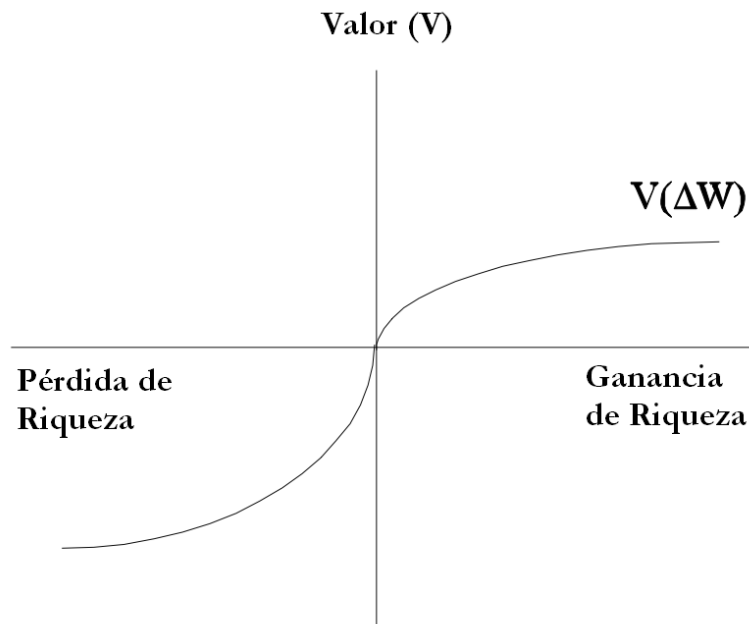


FIGURA 3.1 La función de valoración que es situación-dependiente.
 ΔW es el cambio en la riqueza.
 Nota: el lazo en $\Delta W = 0$ indica aversión a la pérdida.

Una limitación de esta literatura a fecha de hoy, como revela la interpretación de aversión a las pérdidas en el problema de la prima de la acción, es que la dependencia de la situación es una representación incompleta de las preferencias a menos que este unida a una explicación sobre cómo las preferencias se ajustan

dinámicamente a las nuevas situaciones, es decir, cómo evolucionan las preferencias. Los experimentos y otra información presentada aquí muestran que las situaciones inducen las preferencias; pero nos dicen poco sobre el proceso por el cual las personas se adaptan a una situación nueva, ya sean las pérdidas en el mercado de valores, la pérdida de la vista, de la promoción a un cargo de responsabilidad o la transformación de una sociedad sencilla, basada en la caza, a una economía moderna basada en el mercado. Volveré a la evolución de las preferencias bajo la influencia de situaciones económicas que cambian en el capítulo 11.

Las limitaciones y la reformulación consideradas en esta sección han tocado el núcleo formal de la teoría convencional sobre la acción racional. La acumulación reciente de anomalías empíricas en relación al aspecto sustantivo de la teoría, es decir, al axioma sobre el comportamiento basado en el propio interés, también ha motivado las reformulaciones basadas en el concepto de las preferencias sociales.

PREFERENCIAS SOCIALES

En los experimentos sobre el dilema del prisionero a una sola jugada, la tasa de cooperación es por lo general entre un 40 y un 60 por ciento, a pesar de que el abandono de ambos jugadores es la estrategia de equilibrio dominante (Fehr y Fischbacher 2001b). Muchos sujetos prefieren el resultado de la cooperación por ambas partes a la mayor ganancia material que obtendrían si traicionasen a un cooperante. Cuando fallan es porque odian que se les pueda tomar ventaja; muchos abandonan para no arriesgarse a que eso ocurra, no porque sea la estrategia que maximiza sus beneficios independientemente de las acciones de los otros. Estos resultados sugieren que la gente se preocupa por los otros y se preocupan acerca de por qué suceden las cosas independientemente del resultado. Las *preferencias sociales* son las razones del comportamiento *relacionadas con el proceso* y con *las otras personas*.

He aquí un ejemplo sobre las *preferencias relacionadas con el proceso*: usted puede aceptar con ecuanimidad un mal resultado determinado por el lanzamiento de una moneda, mientras que enojadamente rechaza el resultado impuesto por alguien cuya intención sea hacerle daño. Una preferencia que tiene en cuenta el proceso se define como una evaluación basada en las razones de por qué ocurre un estado en lugar de una característica intrínseca del estado. Otros ejemplos incluyen un deseo de ayudar a

los más necesitados sólo si su pobreza es el resultado de la mala suerte y no de la pereza, cumplir promesas y una predisposición a compartir cosas adquiridas por casualidad, pero no las adquiridas por el propio esfuerzo. *El aspecto clave de las preferencias relacionadas con el proceso es que la evaluación del estado es condicional a cómo se llegó a él.* Los comportamientos son sensibles al proceso por dos razones: el proceso que determina un resultado a menudo revela información importante sobre las intenciones de los otros (ej. el pobre que se lo merece) y a menudo dan pistas sobre los comportamientos socialmente apropiados.

Las preferencias basadas en el interés *por los otros* incluyen el rencor, el altruismo y la preocupación por la relación entre las consecuencias para uno y para los demás. Lo que Hobbes llamó el deseo de “eminencia” o una preferencia de resultados “justos” son ejemplos, así como lo es la “emulación pecuniaria” de Thorsten Veblen, ejemplificada por el deseo de “mantenerse bien con los Joneses”(Veblen 1934 [1899]). *El aspecto clave de las preferencias basadas en el interés por los otros es que la evaluación propia de un estado depende de cómo lo experimentan los otros.* En el análisis de las preferencias definidas sobre las experiencias de los otros (así como las propias) sería de ayuda considerar la siguiente taxonomía (ver tabla 3.1) de la distribución de beneficios y costes cuando dos personas interactúan.

TABLA 3.1
Una taxonomía de los comportamientos: costes y beneficios para uno mismo y para los otros

	Coste para sí mismo	Beneficio para sí mismo
Beneficio para los otros	Altruismo	Mutualismo
Coste para los otros	Rencor	Egoísmo

La columna de la izquierda hace una lista de los comportamientos que están específicamente excluidos por el axioma del propio interés. Un comportamiento es *altruista* si confiere un beneficio a otro mientras que inflige un coste para uno mismo (esta definición biológica estándar se restringe a los beneficios y costes y no implica intenciones).Infligir un coste a otro a un coste para uno mismo (abajo a la izquierda) puede estar motivado por rencor, envidia, aversión a la inequidad (si el otro es más rico) o el deseo de castigar a aquellos que le han hecho daño a usted o a otros o que han violado normas sociales. La columna de la derecha es territorio familiar para los

economistas. Dado que en el modelo convencional el cambio de mercado es garantizado por razones de interés por uno mismo, debe conferir beneficios a las dos partes y de ahí que es un ejemplo de lo que los biólogos llaman mutualismo (cuando ocurre entre miembros de diferentes especies). Otros ejemplos incluyen comportamientos aparentemente generosos que incrementan los pagos del individuo en el largo plazo debido a interacciones repetidas o indirectas. Siguiendo a Robert Trivers (1971) estos comportamientos son llamados, algunas veces, “altruismo recíproco”, un nombre inapropiado dado que el altruista recíproco se beneficia de ese comportamiento en cuestión. La terminología del Dalai Lama es más precisa: “la forma estúpida de ser egoísta es...buscar la felicidad solamente para uno...La forma inteligente de ser egoísta es trabajar para el bienestar de los otros” (Dalai Lama 1994:154). Restrinjo el término interés por uno mismo para referirme a los comportamientos de la columna de la derecha y así evitamos el uso tautológico de ese término y hacemos referencia a todo acto que es realizado voluntariamente. El altruista podrá dar obteniendo placer, pero no añadimos claridad alguna si lo llamamos interés por uno mismo.

La observación cotidiana de los otros así como la introspección sugieren que las preferencias por los demás y por los procesos, son importantes. Presentaré de modo abreviado evidencia experimental que confirma estas impresiones. Pero quiero señalar que la evidencia principal para las preferencias sociales no viene de los experimentos sino del mundo económico real y otros comportamientos que son inexplicables en términos de interés por un mismo (sin recurrir a largos razonamientos ad hoc). A algunos de estos comportamientos se ha hecho referencia en este capítulo. Otros ejemplos incluyen los voluntariados en actividades militares peligrosas y otras tareas, el acatamiento de los impuestos por encima de lo que maximizaría los ingresos esperados (en algunos países), la participación en varias formas de acción colectiva y adaptarse a normas y leyes en aquellos casos en los que la transgresión de uno no sería detectada. Los humanos son únicos entre los animales en el grado en el que cooperan con una gran cantidad de no-parientes. Algunas veces esta cooperación es el resultado de instituciones que hacen del comportamiento cooperativo la mejor respuesta para las personas con preferencias de interés por uno mismo (haciendo de la cooperación una forma de mutualismo) pero nadie piensa seriamente que todo ello puede explicarse de esta manera.

Existe una extensa literatura sobre el altruismo, las comparaciones sociales y otros aspectos de las preferencias sociales. Ilustraré la importancia de las preferencias sociales en referencia a la *reciprocidad fuerte*, que no debe confundirse con los comportamientos guiados por el interés por uno mismo descritos por el “altruismo recíproco” de Trivers y con conceptos relacionados como la “reciprocidad indirecta” (confiriendo beneficios a aquellos que han beneficiado a otros y a su vez reciben beneficios como resultado). En contraste a estas “formas inteligentes de ser egoísta”, los motivos de la reciprocidad fuerte pueden inducir comportamientos que son altruistas en el sentido biológico, confiriendo beneficios a otros en el grupo de uno a un coste para sí mismo. Pero la reciprocidad difiere del comportamiento altruista, el que no está condicionado a un tipo de acciones de los otros.

El rechazo de ofertas positivas sustanciales que generalmente se observa en los juegos experimentales del Ultimátum, es un ejemplo de motivos de reciprocidad. Los protocolos experimentales difieren pero la estructura general del Juego del Ultimátum es simple. Los sujetos son puestos en parejas anónimamente para jugar en una sola interacción. Uno es el “que responde” y el otro es el “que propone”. Al que propone se le otorga provisionalmente una suma (“el pastel”, “la olla” u otra metáfora culinaria) que el que responde sabe que debe dividirse entre quien propone y el que responde. El que propone oferta una cierta porción del pastel al que responde. Si el que responde acepta, el que responde obtiene la porción propuesta y quien propone se queda con el resto. Si el que responde rechaza la oferta ninguno obtiene nada. La figura 3.2 presenta una versión del juego en su forma extendida, con los pagos de A primero. En esta versión quien propone simplemente elige entre dos ofertas: dividir el pastel en partes iguales (5,5) o quedarse con 8 y ofrecer 2 al que responde.

En esta situación, el axioma del propio interés predice que las acciones de un individuo son las mejores respuestas definidas sobre los resultados del juego basado en la creencia de que los otros jugadores también satisfacen el axioma del propio interés. Quien Propone A que interesado en sí mismo determinará (por inducción hacia atrás) que el que responde B aceptará la oferta de 2 (porque A cree que B está también interesado en sí mismo) y por eso propondrá la división 8,2, que B aceptará. En juegos en que es posible una oferta más baja que 2, el axioma del propio interés predice que quien propone ofrecerá cero o la suma más baja que sea posible (en la

mayoría de los juegos, quien propone puede elegir cualquier valor en cualquier unidad en que esté denominado el pastel, desde cero hasta el pastel completo).

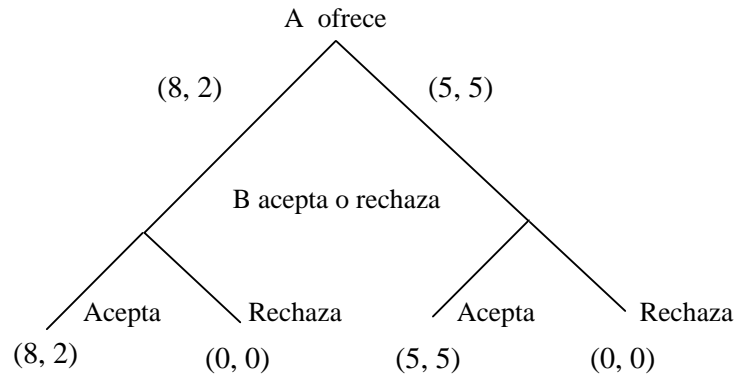


FIGURA 3.2 Un Juego del Ultimátum. Nota: A diferencia de un juego típico, la oferta de A se restringe sólo a $(5,5)$ u $(8,2)$.

Este juego ha sido jugado anónimamente utilizando dinero real en cientos de experimentos con estudiantes universitarios de todo el mundo. La predicción del axioma del interés propio fracasa invariablemente. Las ofertas modales suelen coincidir con la mitad del pastel, en media las ofertas generalmente exceden el 40 por ciento del pastel y ofertas de un cuarto o menores son rechazadas con una probabilidad de entre el 40 y el 60 por ciento. En experimentos llevados a cabo en Estados Unidos, Eslovaquia, Japón, Israel, Eslovenia, Alemania, Rusia, Indonesia, y en otros muchos países, la gran mayoría de quienes proponen ofrecen entre un 40 y un 50 por ciento del pastel (Fehr y Gaechter 2000b).

Estos resultados son interpretados por muchos como una evidencia de la existencia de motivos recíprocos en quien responde, al estar dispuesto a pagar un precio (olvidando un pago positivo) para castigar al que propone por haber realizado una oferta considerada injusta. El comportamiento de quienes proponen es más complicado. Aún el gran número de divisiones iguales (y otras ofertas que parecen justas o casi justas) se explican o por la adhesión a las normas de equidad o altruismo de quien propone o por el propio interés basado en la creencia de que el que responde rechazaría una oferta injusta, aunque ello no puede ser determinado

fácilmente. Ofertas sustanciales violan el axioma del interés propio en cualquier caso, pero quien propone no muestra reciprocidad por la simple razón de que el primer jugador no tiene información sobre B en base a la cual condicionar sus comportamientos. La evidencia que justificaría los motivos de reciprocidad es explicada, por ello, sobre el comportamiento del que responde y no del que propone. Otras interpretaciones – el que responde puede estar buscando implementar un resultado igualitario en lugar de castigar a quien propone, por ejemplo – han sido sugeridas, pero como veremos, la evidencia sobre la existencia de motivos de reciprocidad es bastante convincente.

Los resultados que desafiaron el modelo fundamental de comportamiento en la economía estuvieron sujetos a escrutinio crítico. Algunos se preguntaron si los resultados se debían a que las cantidades en juego eran relativamente bajas. Pero experimentos subsiguientes realizados entre estudiantes universitarios en Indonesia por un “pastel” de un promedio de tres meses de sus gastos reprodujo los resultados del juego (Cameron 1998). Experimentos con estudiantes de Estados Unidos por un pastel de \$100 a \$400 no alteraron los resultados (Hoffman, McCabe, y Smith 1996, Fehr y Fischbacher 2001b). El comportamiento consistente con las preferencias sociales ha sido común en situaciones en las que las cantidades en juego eran elevadas – por ejemplo, un intercambio de regalos en Rusia con ganancias de dos a tres veces el ingreso mensual de los sujetos (Fehr y Fischbacher 2001b). Parece ser que las violaciones a las predicciones del modelo estándar no se deben a que las cantidades en juego sean tan pequeñas que no consigan centrar la atención o provocar los verdaderos motivos de los sujetos experimentales. Otros han sugerido que los sujetos podrían haber malentendido el juego, no obstante experimentos posteriores en los cuales los sujetos jugaron muchas veces con diferentes compañeros no dieron crédito a esta preocupación (Fehr y Fischbacher 2003). Una sugerencia escéptica final fue la de que los sujetos podrían no haber adaptado su comportamiento a la naturaleza no repetitiva de la interacción, tal vez siguiendo las reglas de oro derivadas de interacciones repetidas más comunes. Sin embargo, los sujetos experimentales fácilmente distinguen entre interacciones repetidas y no repetidas (adaptando sus comportamientos como corresponde), y en cualquier caso, el uso de la regla de oro consistente con el juego observado contradice el modelo estándar, sin embargo sucedió. Mientras que el debate acerca de la interpretación de los juegos continúa, existe un consenso sobre la existencia de motivos basados en el interés hacia los otros.

Que los motivos basados en el interés hacia los otros son importantes no es la única lección. Imaginemos que el Juego del Ultimátum en la figura 3.2 es para ser jugado con pequeñas modificaciones en el protocolo. En el experimento llamado Γ_1 , la designación del proponente (ocupado en la figura por A) se determina, como lo es en la mayoría de los experimentos, por el lanzamiento de una moneda: si la moneda dice que A es el proponente el juego es como en la figura 3.2. En Γ_2 el proponente se selecciona como en Γ_1 pero luego se lanza una segunda moneda para determinar que acción tomará A. Entonces A hace la oferta indicada y finalmente B acepta o rechaza. La introspección, así como los resultados experimentales, sugieren que los dos juegos son fundamentalmente diferentes en los comportamientos que provocarán en B, aún cuando B está eligiendo entre pagos idénticos en ambos. En juegos como Γ_2 las ofertas bajas que son aceptadas, serán rechazadas en Γ_1 . Una posible explicación de la diferencia se relaciona con la reciprocidad. En Γ_2 los motivos de reciprocidad no entran en juego porque B sabe que, si el lanzamiento de la moneda dice que la propuesta será de 8,2, A no tuvo la *intención* de hacer una oferta injusta ya que fue simplemente obligado a hacerlo por las reglas del juego. La comparación ilustra las preferencias relacionadas por el proceso: en ambos casos B tiene una oferta mala, pero en el segundo caso el proceso que determina el trato malo es claro que no resulta de las malas intenciones de A. Si los rechazos a las bajas ofertas en Γ_1 fueron motivadas por aversión a la inequidad en lugar de motivos de reciprocidad, por ejemplo, Γ_2 hubiera sido jugado de la misma manera.

TABLA 3.2
Variedades del Juego del Ultimátum

Juegos	Resultados	Interpretación	Fuente
Γ_1 Estándar	Oferta modal $\frac{1}{2}$, ofertas $< 20\%$ rechazadas	Reciprocidad del que responde	Citado en el texto
Γ_2 Ofertas al azar	Pocos rechazos a ofertas bajas	Proponente no responsable	Blount (1995)
Γ_3 Roles elegidos por pruebas	Muchas ofertas bajas, pocos rechazos	Proponente “merecedor”	Hoffman, McCabe, Shachat, y Smith (1994)
Γ_4 "Juego de Intercambio"	Muchas ofertas bajas, pocos rechazos	Marco situacional	Hoffman, McCabe, Shachat, y Smith (1994)
Γ_5 Sin posibilidad de ofertas justas	Ofertas bajas no rechazadas	Las intenciones del proponente importan	Falk, Fehr, y Fischbacher (2003)
Γ_6 Castigo impuesto por un tercero	C castiga ofertas bajas de A a B	Normas de justicia generalizadas	Falk, Fehr, y Fischbacher (2001a)
Γ_7 Estándar : Au/Gnau	Ofertas $> \frac{1}{2}$ son comunes y son rechazadas	Preferencias endógenas y situación – dependientes	Henrich, Bowles, Boyd, Camerer, Fehr, Gintis, McElreath (2001)
Γ_8 Estándar: Machiguenga	Muchas ofertas bajas, muy pocos rechazos	Preferencias endógenas y situación – dependientes	Henrich (2000)

Ahora considere Γ_3 en el cual la posición del proponente no se elige al azar sino sobre la base de una prueba acerca de eventos de actualidad que es tomada con anterioridad a jugar. A, el de mayor puntaje, será el proponente y B quien responde. Los sujetos experimentales juegan Γ_3 de forma diferente que Γ_1 estándar: los proponentes tienen más probabilidad de quedarse con una porción sustancial del pastel y frecuentemente se aceptan propuestas bastante desiguales. Ahora altere el juego una vez más, esta vez simplemente nombrando a Γ_4 “El Juego del Intercambio” en lugar de “Divide \$10”. Como se *llame* el juego no debería tener ningún efecto en el comportamiento de la estructura convencional pero lo tiene: los

proponentes ofrecen menos y se aceptan ofertas menores. Estos y otros experimentos están resumidos en la tabla 3.2.

No es difícil pensar en razones por las cuales las personas juegan Γ_3 diferente a Γ_1 : los que responden pueden sentir que las ofertas menores de los proponentes no deberían ser castigadas dado que reflejan que los proponentes tienen un mayor grado de merecimiento (habiendo ganado el derecho de tener el rol de proponente por su puntaje en la prueba). Pero, ¿Qué pasa con el “Juego del Intercambio”? Parece probable que los sujetos experimentales tomaron el nombre del juego como una pista situacional. Como resultado actuaron más como respuesta a las motivaciones por sí mismo en sus repertorios de comportamiento. Pero aunque uno entienda las diferencias ellos no pueden ser responsabilizados por la estructura de pagos del juego, ya que estos permanecen inalterados por los diversos procesos de asignación de roles, encuadre y la selección de las acciones. Otra variante del juego (Γ_5) reafirma las impresiones acerca de que los rechazos están motivados por el deseo de castigar la injusticia de parte del proponente y no simplemente por el deseo de evitar la aceptación de una división desigual: retiene la oferta de 8,2 del juego estándar pero restringe al proponente a 10,0 (una oferta aún más “injusta”) como la única alternativa a 8,2. Los rechazos a la oferta 8,2 fueron cuatro veces menos frecuentes en Γ_5 que en Γ_1 .

Un rol importante para los valores éticos se sugiere en Γ_6 , el que implica a tres personas y no es estrictamente un Juego del Ultimátum. A asigna una parte del pastel a B (quien simplemente recibe la oferta y no tiene ningún otro rol), luego C, quien ha observado el tamaño del pastel y la oferta puede elegir reducir el pago a A adjudicando parte de la dotación de C (como el pastel que da el experimentador) para ese propósito. Las asignaciones hechas por A de la mitad o más del pastel a B nunca son castigadas, pero cuando A da a B menos de la mitad C está dispuesto a pagar para castigar a A. En este caso, C actúa de forma bastante similar al que responde en el juego del Ultimátum estándar pero está respondiendo a una oferta al parecer injusta, no para el mismo, sino para otra persona (anónima). Fehr y Fischbacher hallaron que el castigo por ese tipo de terceras personas como C es sólo levemente menos duro que el castigo por el que recibe una oferta baja en el montaje del juego del Ultimátum estándar.

También incluyo dos experimentos en los cuales el grupo de sujetos no está – como es usualmente el caso-, compuesto por estudiantes universitarios, sino por miembros de quince sociedades a pequeña escala con poco contacto con los mercados, gobiernos o instituciones modernas. Mis colegas y yo (un equipo de 17 antropólogos y economistas) diseñamos los experimentos para explorar si los resultados de lo que se informa arriba son comunes en sociedades con culturas e instituciones sociales bastante diferentes (Henrich, Bowles, Boyd, Camerer, Fehr, Gintis y McElreath 2004). Las quince sociedades incluían cazadores, pastores y granjeros. Entre la gente Au y Gnao de Papúa, Nueva Guinea, eran comunes las ofertas de más de la mitad del pastel, y las ofertas altas y bajas eran rechazadas con la misma frecuencia. Este resultado, al parecer extraño, no es sorprendente a la luz de la práctica del regalo competitivo dado como un medio de establecer estatus y subordinación en estas y muchas otras sociedades de Nueva Guinea. Contrariamente, entre los Machiguenga en el Perú amazónico, casi tres cuartos de las ofertas eran de un cuarto del pastel o menos y había solamente un rechazo, un patrón sorprendentemente distinto al de los experimentos conducidos hasta ese momento. Sin embargo, aún entre los Machiguenga la oferta media era de 27 por ciento, sugiriendo que las ofertas superaron la oferta esperada de maximización de pagos.

El análisis del experimento en las quince sociedades que estudiamos nos llevó a las siguientes conclusiones: Los comportamientos son altamente variables entre grupos, ni un solo grupo aproximó los comportamientos implicados por el axioma del interés propio, y entre grupos, las diferencias en los comportamientos parecen reflejar diferencias en los tipos de interacción social experimentados en la vida cotidiana. La evidencia de las condiciones económicas que afectan las normas del comportamiento es bastante convincente. Por ejemplo, los Aché en Paraguay comparten en partes iguales entre todos los miembros del grupo algunos tipos de comida (carne y miel) adquirida a través de la caza y la recolección. La mayoría de los proponentes Aché contribuyeron con la mitad del pastel o más. De la misma forma, entre los cazadores de ballenas Lamalera de Indonesia, quienes cazan en grandes equipos y dividen su caza de acuerdo con normas estrictas de repartición, la propuesta promedio fue de 58% del pastel. Además, los cazadores de ballenas jugaron un juego de manera muy diferente al de los estudiantes universitarios de Indonesia que mencionamos arriba.

El juego del Ultimátum es uno de muchos en los que los sujetos experimentales se comportaron de formas que discrepan mucho con las predicciones del axioma del interés propio. Colin Camerer y Ernst Fehr (2004) estudiaron siete juegos en los que los experimentos sugirieron la supremacía de las preferencias sociales. Uno de estos, el Juego de los Bienes Públicos, es importante como una analogía a muchos problemas económicos mundiales e instructivo sobre el comportamiento humano. Algunas veces se lo llama dilema del prisionero de n – personas porque tiene la misma estructura de incentivos: si los jugadores se adaptan al axioma del interés propio sin contribuir en nada al bien público (análogo a la desertión) es el equilibrio de estrategia dominante, pero la contribución universal maximiza los pagos totales. He aquí el juego: se les da a n jugadores una “dotación” y para luego, simultáneamente seleccionar una suma $c_i \in [0, Y]$ para contribuir al bien público. Los pagos a cada jugador son $\pi_i = y - c_i + m \sum_j c_j$ para $j = 1 \dots n$. Esto describe un Juego del Bien Público si $m < 1 < mn$. La primera de estas desigualdades implica que la mejor respuesta del individuo es contribuir con nada y la segunda implica que los pagos totales (sumados los del grupo) se maximizan si todos contribuyen con su dotación entera. Variantes del Juego del Bien Público han sido usadas para mostrar la extracción individual de un recurso ambiental de fondo común; las explicaciones incluyen contribuciones a proyectos conjuntos como el pago de impuestos y la participación en huelgas.

La predicción del axioma del interés propio ($c_i = 0 \ \forall i$) se contradice constantemente en los experimentos (examinada por Ledyard 1995). En juegos de una interacción las contribuciones promedian alrededor de la mitad de la dotación mientras que en juegos de múltiples interacciones las contribuciones comienzan alrededor de la mitad y luego descienden, de manera que la mayoría de los jugadores contribuyen con nada en la ronda final de un juego a 10 rondas. Se pensó inicialmente que este descenso en las contribuciones confirmaba el modelo convencional, la idea era que una vez los sujetos han entendido el juego no contribuyen con nada. Pero un experimento en el cual se comenzó inesperadamente un segundo juego de bienes públicos de diez rondas al finalizar el primer juego de 10 rondas, sugiere que este no es el caso: en el segundo juego los jugadores otra vez comenzaron con contribuciones de la mitad. Muchos interpretaron el descenso en las contribuciones como un reflejo de la desilusión acerca de las expectativas que otros contribuirían con más, junto con el deseo de castigar a los que contribuían con poco

(o, al menos, de no ser aventajados) en una situación en la que esto sólo podría hacerse reduciendo las contribuciones propias.

Un fuerte apoyo a este último enfoque es provisto por el ingenioso experimento diseñado por Fehr y Gaechter (2002): tiene la misma estructura del bien público de arriba excepto que luego de las contribuciones de los individuos, las contribuciones se hicieron conocidas (sólo por un número de identificación, no por nombre, por supuesto) a todos los miembros del grupo quienes luego tuvieron la oportunidad de castigar a otros del grupo, imponiendo un costo (reducción del pago) tanto al castigador como al castigado.⁹ En un experimento con este juego, Fehr y Gaechter adoptaron lo que se llama tratamiento entre perfectos extraños: después de cada ronda del experimento a diez rondas los grupos fueron mezclados otra vez por lo que los jugadores sabían que ningún jugador se encontraría con otro jugador más de una vez. Así, la motivación para el castigo no podía ser el interés propio. Si los que contribuían con poco responderían al castigo contribuyendo con más en rondas posteriores los pagos de otros subían, pero no el del castigador (debido al tratamiento como perfectos extraños). Así, el castigo no es más un bien público que la contribución inicial. Este es el caso sobre la última ronda del juego cuando la última acción tomada por cualquier jugador es la de comprometerse con un castigo costoso de los miembros compañeros del grupo: aquellos que castigan bajo estas condiciones deben valorar el castigo per se en lugar de cualquier consecuencia anticipada de castigo de los pagos de su juego basado en la modificación de los comportamientos de los otros.

En el juego del Bien Público con castigo de Fehr y Gaechter, las contribuciones comenzaron por alrededor de la mitad de la dotación (como en el juego estándar) pero luego se *elevaron* en lugar de caer a lo largo del juego. Mis coautores y yo (de lo que se informó en Bowles y Gintis 2002b) implementamos un juego similar en que confirmamos lo que uno esperaría: el castigo se dirige a los bajos contribuyentes y ellos responden fuertemente al castigo. Aquellos que pensaron que podían hacer trampa en la última ronda del juego reduciendo sus contribuciones pagaron caro por su error. También hallamos algo bastante inesperado. Cuando aquellos que contribuyeron *por encima* de la media fueron castigados (como ellos lo habían sido ocasionalmente) *redujeron* abruptamente sus contribuciones. Aún más

⁹ Un experimento anterior de este tipo con resultados similares es el de Ostrom, Gardner y Walker (1994).

sorprendente es el hecho que la respuesta positiva al castigo de los bajos contribuyentes no fue una mejor respuesta definida sobre los pagos del juego. Tomando en cuenta lo observado, la relación entre la cantidad de castigo esperada y la oferta propia, una contribución de cero siguió siendo la mejor respuesta, pero, sin embargo, aquellos castigados respondieron contribuyendo con más.

Una interpretación razonable de estos experimentos es que, como en el Juego del Ultimátum, la gente está dispuesta a pagar para castigar a aquellos que violan normas sociales aún cuando no hay expectativa de pagos futuros o indirectos. En otras palabras, los sujetos estaban actuando de acuerdo con motivos de reciprocidad. Pero algo más parece estar actuando. El hecho que el castigo indujo más contribuciones de los evasores (al contrario de la opción de maximización de los pagos, aún cuando el supuesto castigo es tenido en cuenta) sugiere que la sanción social por los pares puede movilizar sentimientos de vergüenza en situaciones en las que el castigo conlleva alguna legitimidad (a los ojos de la persona castigada). En dos experimentos similares – uno de laboratorio y otro de campo entre granjeros en Zimbabwe – el “castigo” condujo solamente a un sentimiento de disgusto y no redujo los pagos del castigado. Pero el hecho que aquellos que fueron castigados contribuyeron más en períodos subsiguientes muestra los fuertes efectos de la sanción social consistente con la interpretación de la “vergüenza” (Barr 2001, Masclet, Noussair, Tucker y Villeval 2003). En el capítulo 4 presento un modelo de cómo las preferencias sociales como la vergüenza y la reciprocidad pueden apoyar la cooperación en las interacciones de bien público.

El Juego del Bien Público provee un buen ejemplo de comportamientos y estructura situación -dependientes. Jean Ensminger condujo un experimento de bien público con los Orma, un pueblo de pastores en Kenia, como parte de un proyecto experimental multicultural mencionado anteriormente. Cuando los Orma necesitan algún bien público – una nueva escuela primaria o la reparación de una carretera, por ejemplo – se le pide a los miembros de la comunidad que voluntariamente contribuyan al proyecto con cantidades crecientes en la suma de riqueza (ganado) de la familia. Este sistema de provisión voluntaria de bienes comunes se llama *harambee*. Cuando Ensminger explicó a los sujetos el Juego de los Bienes Públicos, ellos inmediatamente lo llamaron el “Juego Harambee” y sus contribuciones fueron predichas por su riqueza (en el mundo real), igual que hubiese sido el caso en un Harambee real. Cuando los sujetos Orma jugaron el Juego del Ultimátum no hicieron

la analogía con el harambee (o aparentemente con ningún otro aspecto de su vida cotidiana) y la riqueza no predijo ningún aspecto de su juego experimental.

¿Las personas se comportan en escenarios naturales de la misma forma que lo hacen en los experimentos? La relación de los comportamientos entre el juego experimental y el mundo real es compleja y no quiero afirmar una correspondencia demasiado cercana entre los dos. Contrario a las esperanzas (desacertadas, en mi opinión) de algunos experimentadores, los juegos experimentales no mantienen los motivos abstractos sin contaminar por las situaciones. En esto, el juego experimental se parece a cualquier otro comportamiento y el experimento es sólo otra situación.¹⁰ La situación del juego, las instrucciones del experimentador y el gusto son un marco muy fuerte y no podemos esperar que no tengan ningún efecto. Los experimentos no revelan una esencia de la naturaleza humana universal. Simplemente muestran que los comportamientos comunes en interacciones sociales genéricas se explican claramente por las preferencias sociales, lo que sugiere que muchos de los ejemplos del mundo real de aparentes violaciones del axioma del interés propio no son el resultado de peculiaridades de los ejemplos particulares del mundo real.

UNA FUNCIÓN DE PREFERENCIAS SOCIALES CON BASES EMPÍRICAS

En respuesta a las violaciones al axioma del interés propio en un número de experimentos, los economistas han intentado reformular una función de utilidad capaz de explicar los comportamientos de una manera parsimoniosa. ¿Existe una función de utilidad que sea lo suficientemente simple para ser rastreada y lo suficientemente fuerte para explicar no sólo una de las anomalías experimentales sino todas ellas? En este momento existe un número de funciones de utilidad que son capaces de explicar un gran rango de comportamientos experimentales (Falk y Fischbacher 1998, Fehr y Schmidt 1999, Bolton y Ockenfels 1999, Rabin 1993, Charness y Rabin 1999, Levine 1998). Los ingredientes básicos de la función de utilidad propuesta son el interés propio, altruismo, rencor, mentalidad justa y

¹⁰ Lowenstein (1999) da una evaluación escéptica pero balanceada. Los comportamientos en los juegos han sido mostrados para predecir comportamientos de la vida real en algunos pocos casos: aquellos que confiaron en un experimento de confianza por Glaeser, Laibson, Scheinkman y Soutter (2000), por ejemplo, mostraron más confianza en un número de situaciones del mundo real. Contrario a esto, las respuestas a preguntas de una encuesta estándar sobre confianza no tuvieron ninguna correlación con ningún comportamiento medido (experimental o no experimental).

reciprocidad. Las funciones difieren en la forma en que se combinan esos componentes y los tipos de comportamiento que los autores quieren destacar.

He aquí una función de utilidad (propuesta por Fehr y Schmidt) que toma en cuenta tanto el interés propio como lo que ellos llaman “aversión a la inequidad.” Una función de utilidad justa (esto es, con aversión a la inequidad) de la persona i (interactuando sólo con otra persona, j) está dada por:

$$U_i = \pi_i - \delta_i \max(\pi_j - \pi_i, 0) - \alpha_i \max(\pi_i - \pi_j, 0) \quad (3.3)$$

donde π_j y π_i son los pagos materiales de los dos individuos y $\delta_i \geq \alpha_i$ y $\alpha_i \in [0,1]$. Esta función de utilidad expresa la valoración individual de i de su propio pago así como su aversión a la diferencia de pagos, con diferencias desventajosas ($\pi_j - \pi_i > 0$) siendo sopesadas más fuertemente (δ_i) que las diferencias ventajosas (α_i). El límite más alto sobre α excluye lo que podría llamarse niveles de “auto castigo” de aversión a las inequidades ventajosas: un individuo con $\alpha = 1$ se preocupa solamente por los pagos de los otros (si no le alcanza con lo propio). Contrario a esto, una persona (i) muy aversa a la inequidad desventajosa puede preferir $\pi_i = \pi_j = 0$ a $\pi_i = 1$ y $\pi_j = 2$, por lo que δ puede exceder de 1.

Para ver las implicaciones de la mentalidad justa tanto para las conductas de compartir y las de castigar, imagine que los dos deben dividir una unidad ($\pi_i + \pi_j = 1$) y que $\alpha_i > 1/2$. En este caso, $dU_i / d\pi_i < 0$ para todas las divisiones tales que $\pi_i - \pi_j > 0$. Por lo que la preferencia de compartir del individuo i será dividir la unidad en partes iguales (aún cuando la división inicial favorecía a i por sobre j , i preferirá transferir algo de su pago a j). De la misma manera, si $\delta_i \geq 1/2$ y los pagos se dividieran para que j fuera a recibir 0.6 e i 0.4, i estará dispuesto a pagar 0.1 para reducir los pagos de j en 0.3 para que ambos reciban 0.3. Aún más sorprendente, en este caso i rechazaría una oferta de menos de 0.25 si haciéndolo ambos no recibirían nada (como en el Juego del Ultimátum).

La mentalidad justa puede explicar otra anomalía experimental mencionada al comienzo: un número sustancial de sujetos experimentales en juegos del dilema del prisionero de una interacción cooperan (a pesar que la deserción es la estrategia dominante en los pagos del juego). Un jugador de la fila con mentalidad justa (uno con la función de utilidad Fehr – Schmidt de más arriba) enfrentando los pagos

materiales del dilema del prisionero estándar $a > b > c > d$ cooperaría si sabe que el jugador de la columna cooperará siempre y cuando la desutilidad que experimentó de la desigualdad ventajosa sea suficientemente grande, o $\alpha > (a - b) / (a - d)$ (Ver tabla 3.3).

TABLA 3.3
Dilema del Prisionero Estándar y Pagos de Utilidad
un jugador de las filas con mentalidad justa

	Cooperar	Desertar
Cooperar	b	d
	b	d - $\delta(a - d)$
Desertar	a	c
	a - $\alpha(a - d)$	c

Nota: Los pagos de utilidad para el jugador de mentalidad justa de las fila están en negrita.

Si se obtiene esta desigualdad (lo que puede pasar dado que el lado derecho es necesariamente menor que la unidad), entonces el juego resultante ya no es el dilema del prisionero sino el Juego del Aseguramiento, dado que existe algún valor crítico de $p^* \in (0,1)$ tal que si Fila creyera que Columna desertará con una probabilidad menor a p^* , entonces su mejor respuesta es cooperar. También se puede mostrar claramente que $dp^* / d\alpha > 0$ mientras que $dp^* / d\delta < 0$, por lo que si esta interacción toma lugar entre jugadores de mentalidad justa en parejas elegidas al azar, en un ambiente evolucionista del tipo que se usó como modelo en el capítulo anterior, agrandar la desutilidad de la inequidad ventajosa aumenta la cuenca de atracción del equilibrio de la cooperación mutua mientras que incrementar la inutilidad desigual desventajosa hace lo opuesto.

En un experimento diseñado para estimar los parámetros de una función como la ecuación (3.3) Lowenstein, Thompson y Bazerman (1989) crearon una variedad de escenarios que tenían en común que una suma debía ser dividida, pero las situaciones diferían en la relación personal de los participantes (negativa, neutral o positiva) y en la naturaleza de la interacción (comercial, otra). Encontraron que la desigualdad desventajosa disgustaba fuertemente, sin importar la naturaleza de la relación personal o de la transacción. Contrario a esto, la desigualdad ventajosa disgustaba al 58 por ciento de los sujetos en las transacciones que no eran comerciales, pero era

preferida por la mayoría en las transacciones de comerciales, causando disgusto sólo al 27 por ciento. La naturaleza de la relación personal también importó: en un ambiente de relaciones personales neutras o positivas al 53 por ciento le disgustó la desigualdad ventajosa, mientras que en un ambiente de relaciones personales negativas sólo lo hizo con el 36 por ciento. Este experimento muestra evidencia directa de la aversión a la inequidad y también es consistente con el punto de vista según el cual los comportamientos están usualmente condicionados por las creencias propias acerca de la otra persona (positiva o negativa) y son específicas a cada situación (comerciales o no).

Las preferencias de mentalidades justas se definen sobre los resultados, pero las preferencias recíprocas dependen también de la creencia propia sobre de la intención o el tipo de individuo con el que se trata. Siguiendo ideas inicialmente esgrimidas por Rabin (1993) y Levine (1998) la siguiente función incorpora el interés propio, altruismo y reciprocidad. La utilidad de un individuo depende de su propio pago material y el de otros individuos $j = 1 \dots n$ de acuerdo con

$$U_i = \pi_i + \sum_j \beta_{ij} \pi_j \text{ para } i \neq j \quad (3.4)$$

donde β_{ij} , el peso del pago material de j en las preferencias de i .

$$\beta_{ij} = \frac{a_i + \lambda_i a_j}{1 + \lambda_i} \quad \forall j \neq i \quad (3.5)$$

y $a_i \in [-1, 1]$ y $\lambda_i \geq 0$. El parámetro a_i es el nivel de buenos deseos incondicionales o de malos deseos (altruismo o rencor) de i hacia los demás, y $a_j \in [-1, 1]$ la creencia de i con respecto a los buenos deseos de j mientras que λ_j indica cuanto condiciona i su evaluación de los pagos de otros en (creencias sobre) el tipo del otro. Si $a_i = 0$ y $\lambda_j > 0$, entonces el individuo i actúa con reciprocidad no altruista (no muestra ni buenos deseos ni rencor incondicionales pero condiciona su comportamiento en la bondad o el rencor de otros).

Si $\lambda_j = 0$ y $a_j \neq 0$, entonces i muestra un altruismo incondicional o rencor dependiendo del signo de a_j . El denominador es aumentado por λ_j por lo que $\beta_{ij} \leq 1$, de tal modo que restringe la valoración que uno hace de los pagos de los demás para que no sean mayores a los propios. Note que $d\beta_{ij}/d\lambda_j$ tiene el signo de $(a_j - a_i)$,

lo que significa que el nivel de reciprocidad afecta la extensión en la que los pagos de los otros entran en la evaluación propia incrementándola si el otro es más bondadoso que uno mismo, y a la inversa. Si $a_j = a_i$ entonces $\beta_{ij} = a_i$ para cualquier nivel de reciprocidad.

Al igual que la función de aversión a la inequidad esta función de utilidad basada en la reciprocidad puede usarse para explicar comportamientos generosos y castigadores. Sin embargo, el análisis es considerablemente más complicado. En la mayoría de las interacciones sociales tenemos creencias previas acerca de los tipos de otros basadas en el conocimiento de su comportamiento anterior, pistas basadas en otros hechos sobre ellos (incluyendo su estatus como un “miembro” (*insider*) o como un “no miembro” (*outsider*) en la interacción actual) y la situación misma. Así, las creencias propias acerca de los tipos de otros y de allí la evaluación que uno hace de sus beneficios depende plausiblemente de las acciones pasadas, lo que depende de sus creencias en el tipo propio, y así sucesivamente. Si uno actúa con reciprocidad y cree que los otros son altruistas, uno podría participar con generosidad condicional. Pero si la generosidad no es correspondida uno puede actualizar sus creencias acerca de los tipos de los otros y participar en el castigo o, al menos, retractar la generosidad, como se hizo visible en los experimentos de bienes públicos. Así, puede haber comportamientos tanto patrón –dependientes como específicos a la situación: una situación que induce creencias acerca que los otros son altruistas puede mantener niveles de generosidad altos y sostenibles, mientras que los mismos individuos interactuando en otra situación pueden participar en castigos de rencor mutuamente costosos. La naturaleza de los comportamientos específica a las situaciones o la dependencia de las trayectorias podría explicar porqué las jugadas de los individuos se ven tan afectadas por cambios en los protocolos experimentales que serían irrelevantes si los modelos convencionales estuvieran correctos. También puede aclarar por qué esas diferencias grandes en los comportamientos se encuentran en nuestro estudio intercultural.

CONCLUSIÓN

Las funciones de aversión a la inequidad y las basadas en la reciprocidad que se acaban de presentar, son pasos importantes hacia la construcción de una concepción del comportamiento más adecuada. Pero el proceso sigue y está lejos de completarse.

La evidencia sobre que la aversión a la inequidad y los motivos de reciprocidad son comunes no sugiere que las personas sean irracionales. En realidad, evidencia experimental sólida indica que cuando los individuos dan a otros (ej. en el Juego del Dictador) su comportamiento se adapta a los supuestos de transitividad y a otros requerimientos de la elección racional (Andreoni y Miller 2002). Además, las personas responden al precio de dar, dando más cuando les cuesta menos beneficiar a los demás. Así, la importancia de los motivos del interés por el otro no desafían las presunciones de racionalidad sino que sugieren que los argumentos de la función de utilidad deben expandirse para explicar las preocupaciones de los individuos hacia los demás.

La evidencia experimental, entre otras, también sugiere que una formulación adecuada debe tener en cuenta la heterogeneidad de comportamientos de la mayoría de los grupos humanos. Usando la información de una amplia gama de experimentos, Ernst Fehr y Simon Gächter estiman que entre el 40 y el 66 por ciento de los sujetos muestran elecciones recíprocas. Los mismos estudios sugieren que entre el 20 y el 30 por ciento de los sujetos muestran preferencias convencionales de interés propio orientadas a los resultados (Fehr y Gächter 2000b, Camerer 2003). Loewenstein, Thompson y Bazerman (1989) distinguieron en sus experimentos entre los siguientes tipos:

Los *Santos* prefieren consistentemente la igualdad y no les gusta recibir pagos mayores que las otras partes aún cuando tienen una relación negativa con el oponente...a los *leales* no les gusta recibir pagos mayores en relaciones neutras o positivas, pero buscan la desigualdad ventajosa en relaciones negativas... los *competidores despiadados* prefieren, de forma consistentemente, salir ganando sin importar el tipo de relación. (p.433)

De sus sujetos, el 22 por ciento eran santos, el 39 por ciento eran leales y el 29 por ciento eran competidores despiadados (el resto no pudo ser clasificado).

Así, el objetivo de una reformulación de las bases del comportamiento en la economía no debe ser un nuevo *Homo sociologicus* que reemplace al *Homo economicus* sino un marco capaz de tomar en cuenta la heterogeneidad. Esta tarea es esencial dado que la heterogeneidad hace la diferencia en los resultados, pero es desafiante porque los efectos no son captados adecuadamente por un proceso de simple promedio. El resultado de la interacción entre una población que está compuesta de igual número de santos y de competidores despiadados no será generalmente el

promedio de los resultados de dos poblaciones con sólo un tipo porque las pequeñas diferencias en la distribución de los tipos en una población puede tener grandes efectos en cómo se comporta *cada uno*.

Además, diferencias que parecen pequeñas en instituciones pueden hacer grandes diferencias en los resultados. Imagine un juego del Dilema del Prisionero de una interacción que se juega entre un jugador de interés propio (para quien Desertar es la estrategia dominante en el juego de jugadas simultáneas) y uno que actúa con reciprocidad (quien prefiere Cooperar si el otro coopera y en caso contrario Desertar) (Fehr y Fischbacher 2001b). Imagine que los tipos de los jugadores son conocidos por cada uno. Si el juego se juega simultáneamente el que actúa con reciprocidad, a sabiendas que el otro va a Desertar, hará lo mismo. El resultado será de desertación mutua. Si el jugador de interés propio mueve primero, sabrá que el que actúa recíprocamente hará lo mismo cualquiera sea la acción que tome, acotando los resultados posibles a {Cooperar, Cooperar} o {Desertar, Desertar}. El jugador interesado por sí mismo entonces cooperará y la cooperación mutua será tenida por resultado. Recuerde como otro ejemplo que en el Juego de Bienes Públicos con Castigo, aquellos con preferencias de reciprocidad no sólo actuaban generosamente, sino que aparentemente inducían a los tipos egoístas a actuar *como si* fueran generosos. Pero si hubiera habido muy pocos que actúan con reciprocidad, todos los jugadores (tanto los del tipo que actúan recíprocamente como los del interés en sí mismo) hubieran convergido a una contribución de cero.

Además de la heterogeneidad entre los individuos, la versatilidad de los mismos debe tenerse en cuenta. En el Juego del Ultimátum los proponentes a menudo ofrecen sumas que maximizan sus pagos esperados dada la relación observada entre ofertas y rechazos: ellos se comportan interesados en sí mismos *pero esperan que los que responden no lo hagan*. Además, los *mismos individuos* cuando están en el rol del que responde típicamente rechazan ofertas sustanciales si parecen ser injustas, confirmando así las expectativas del proponente y violando el axioma del interés propio.

Finalmente, como hemos dicho anteriormente (y discutiremos en el capítulo 11), las preferencias son de alguna manera aprendidas en lugar de estar dadas exógenamente: cambios duraderos en las razones para el comportamiento de un individuo a menudo suceden como resultado de la experiencia propia. Esto significa

que poblaciones que experimentaron diferentes estructuras de interacción social durante períodos prolongados son propensas a exhibir comportamientos diversos, no simplemente porque los límites e incentivos que conllevan estas instituciones son diferentes sino también porque la estructura de la interacción social afecta la evolución de ambos repertorios de comportamiento, la forma en que las situaciones inducen comportamientos y la forma en que son evaluados los resultados. (Porque el funcionamiento de las instituciones depende de las preferencias de los individuos involucrados, también será el caso en que las instituciones sean endógenas con respecto a las preferencias; yo pongo como ejemplo el proceso resultante, llamado la *coevolución de las preferencias e instituciones*, en los capítulos 11 al 13.)

El progreso en dirección a una base del comportamiento más adecuada debe tomar en cuenta estos tres aspectos de las personas: *su heterogeneidad, versatilidad y plasticidad*.

Las nuevas teorías también deben abordar dos desafíos. El primero concierne al estatus normativo de las preferencias. Si las preferencias explican comportamientos no pueden hacer el trabajo de evaluar resultados sin asistencia. Los motivos son que algunas razones comunes para el comportamiento – debilidad de la voluntad, rencor y adicción vienen a la mente – generalmente inducen comportamientos cuyos resultados pocos tolerarán.

El segundo desafío aparece porque la evidencia experimental y otra que indica la importancia de las preferencias sociales genera un interrogante evolutivo difícil. Si muchos de nosotros somos de mentalidad justa y recíproca, entonces debimos haber adquirido estas preferencias de alguna manera, y sería una buena verificación de la plausibilidad de las teorías de la preferencia social y la evidencia empírica en la que están basadas para ver si se puede dar una explicación razonable al éxito evolutivo de estas preferencias. La generosidad con los parientes genéticos se explica claramente. El interrogante evolutivo tiene que ver con comportamientos no egoístas con no parientes (lo que significa comportamientos con costos individuales sin beneficio o la columna izquierda en la tabla 3.1 de arriba). Entre no parientes, las preferencias egoístas parecen ser favorecidas por cualquier proceso evolutivo monotónico de pagos, tanto genéticos o culturales. Así, la mentalidad justa que induce a las personas a transferir recursos a los menos favorecidos y los motivos de reciprocidad que nos obligan a incurrir en costos de castigar a aquellos que violan las normas del grupo,

están condenados a la extinción por procesos evolutivos a largo plazo. Si las preferencias sociales son comunes, esta explicación evolutiva convencional debe ser incorrecta.

En próximos capítulos retomo esta pregunta y doy una serie de modelos que explican el éxito evolutivo de las preferencias sociales. Particularmente, exploraré la contribución al éxito evolutivo de rasgos no egoístas hecha por estructuras características de interacción social humana, a saber, segmentación social, interacciones repetidas y construcción de reputación (en el capítulo 7) y el cumplimiento de normas a nivel del grupo y conflictos entre grupos (en los capítulos 11 y 13). En muchos casos el éxito evolutivo de lo que parecen ser rasgos no egoístas se explica por el hecho que cuando se toman en cuenta los efectos indirectos y a largo plazo, los comportamientos son de maximización de los pagos, usualmente representando formas de mutualismo. Pero también presentaré modelos plausibles para explicar el éxito evolutivo de comportamientos que benefician a otros miembros del grupo propio con un costo para sí mismo.

Como la teoría de las preferencias sociales, la teoría de la prospectiva también trae interrogantes evolutivos. Los agentes con descuento hiperbólico actúan de forma inconsistente en el tiempo, sus pagos promedio a través de un largo período se incrementarán si se adaptan a los dictados del modelo de utilidad descontada. De la misma manera aquellos que sopesan eventos de baja probabilidad recibirán pagos esperados más bajos que los competidores que si realizan una maximización apropiada de la utilidad esperada. Esto no significa que aquellos que usan el descuento de manera inconsistente en el tiempo y violan los axiomas de la utilidad esperada están condenados, pero dado que la evolución genética o cultural tiende a favorecer a aquellos con pagos más altos, si plantea un interrogante. De la misma forma, los individuos aversos a las pérdidas preceden oportunidades de ganancias sustanciales esperadas en situaciones riesgosas. Su aversión a las pérdidas los pone en desventaja en competencia con otros a quienes su función de utilidad no está ligada al status quo. Estos enigmas evolutivos que nacen de la teoría prospectiva han recibido menos atención que el interrogante de las preferencias sociales. No me referiré a ellos en extenso excepto para hacer notar que la evidencia inicial del

descuento hiperbólico viene de pichones y ratas, por lo que esto no es únicamente un comportamiento humano.¹¹

En el capítulo 4 generalizo los tipos de problemas de coordinación introducidos en el capítulo 1 como juegos de 2 X 2 y analizo la impresionante variedad de instituciones, normas y otras formas que han desarrollado las personas para evitar o atenuar las fallas de coordinación. Las preferencias sociales, como veremos, juegan un papel central en este proceso.

¹¹ El descuento hiperbólico en humanos y otros animales se describe en Ainslie (1975). Green y Myerson (1996) y Richards, Mutchell, de Wit y Seiden (1997).