

Racionalidad, Economía e Interdisciplinariedad

Compilador: Gustavo Marqués

Staff

Director
Gustavo Marqués

Coordinador Técnico
Hernán Ruggeri

Secretaría
Analía Cocco

Editor Responsable

Centro de Investigación en Epistemología de las Ciencias
Económicas

Facultad de Ciencias Económicas
Universidad de Buenos Aires

Av. Córdoba 2122 1º p. Aula 111
(1120) Ciudad Autónoma de Buenos Aires
Argentina
Tel. (54-11) 4370-6152
Correo electrónico: ciece@econ.uba.ar

ISBN: 978-950-29-1182-3

Esta publicación se terminó de compaginar e imprimir en el mes de

Noviembre de 2009

Queda hecho el depósito que establece la ley 11.723

No se permite la reproducción total o parcial de este libro, ni su almacenamiento en un sistema informático, ni su transmisión en cualquier forma o medio electrónico, mecánico u otros medios, sin permiso del editor.

ÍNDICE

Introducción.....	7
Primer Sección: Análisis histórico-conceptual de la noción de “utilidad”.....	9
La Evolución De La Noción De Valor Económico Y La Teoría De La Decisión, <i>por</i> Eduardo Scarano.....	11
La Utilidad Como Cualidad Hedónica: Sus Avances, Limitaciones, Y Principales Campos De Investigación Aplicada, <i>por</i> Leonardo Ivarola.....	59
Segunda Sección: Crítica de las nociones convencionales de racionalidad y utilidad.....	103
De La Neurología A La Decisión Racional: El Camino Del Funcionalismo Homuncular, <i>por</i> Alejandro Miroli.....	105
¿De Qué Sirve Maximizar La Utilidad Esperada?, <i>por</i> Gustavo Marqués.....	173
Sobre La Inutilidad De Las Encuestas, <i>por</i> Diego Weisman.....	215
Tercer Sección:	
Neurociencia e Interdisciplinariedad.....	247
The Possibilities Of Neuroeconomics: An Account Through The Lens Of Economic Methodology, <i>por</i> Roberta Muramatsu.....	249

Neuroeconomics: Issues Of Scientificity And Interdisciplinarity
Por Uskali Maki.....299

INTRODUCCIÓN

La teoría de la elección racional y sus aplicaciones en economía han sido el centro de fuertes controversias. En los últimos años éstos debates se han intensificado debido a la influencia de diversas corrientes de pensamiento que se apartan en mayor o menor medida de las visiones usuales acerca del comportamiento racional (maximizador) en economía. Ya a comienzos de los años 50 comenzó a advertirse que la concepción convencional de la elección racional bajo condiciones de riesgo se aplicaba imperfectamente al ámbito de las decisiones humanas, lo que dio lugar al desarrollo de teorías alternativas que impulsaban debilitamientos del concepto usual de racionalidad, las cuales fueron agrupadas bajo los rótulos de Racionalidad Acotada (Bounded Rationality) y Racionalidad Procedimental (Procedural Rationality).

Inspirándose en estas objeciones y, posteriormente, en trabajos experimentales (y teóricos) de psicólogos y economistas, se desarrolló un vasto movimiento de pensamiento económico, conocido como Economía Conductual (Behavioral Economics). Y, como suele suceder

cada vez que los usos establecidos se resquebrajan y se amplía la mirada, un conjunto creciente de biólogos y economistas se ha sumado al debate produciendo un abundante material de investigación al que se designa como Neuro-economía (Neuroeconomics).

Los trabajos incluidos en esta compilación reflejan una diversidad de posturas vinculadas con estas problemáticas. Se trata de una selección de trabajos expuestos durante el mes de octubre de 2008 en el marco de una sección especial de las XIV Jornadas de Epistemología de las Ciencias Económicas, denominada “Evaluación metodológica de la teoría económica estándar de la decisión racional y sus relaciones con las disciplinas colindantes, la psicología y la neurobiología”. El evento que tuvo lugar en la Facultad de ciencias Económicas de la Universidad de Buenos Aires.

Primer Sección:

Análisis histórico-conceptual de la noción de “utilidad”

LA EVOLUCIÓN DE LA NOCIÓN DE VALOR ECONÓMICO Y LA TEORÍA DE LA DECISIÓN

Eduardo R. Scarano (CIECE, FCE-UBA)

I. Introducción

En el presente artículo se expone la evolución del valor económico, una de las nociones que conducen a la formulación de la teoría de la decisión. Esta teoría suele también denominarse teoría de la racionalidad; un agente actúa racionalmente, simplemente es racional, si satisface los axiomas que la caracterizan. La noción de racionalidad tiene una venerable tradición que comienza con la Filosofía misma. Platón y Aristóteles, por solo citar dos filósofos antiguos, desarrollaron sus propias concepciones de la racionalidad. Pero, así como el atomismo tuvo su origen con Demócrito en la filosofía griega y recién con la física moderna tuvo un desarrollo científico que perdura hasta nuestros días; de la misma manera, es con la economía que la racionalidad adquiere un desarrollo científico, en particular, con el marginalismo primero y luego con la economía neoclásica.

Si a la Filosofía debemos el origen del concepto, y a través del tiempo la incorporación de novedosas nociones acerca de la razón y de la racionalidad, a la economía debemos un desarrollo en profundidad, sistematicidad, poder empírico y explicativo incomparable. Esta dirección se puede ilustrar con el modelo de decisión bajo riesgo e incertidumbre que invirtió los términos con la filosofía y ahora sirve a los filósofos para discutirla, examinar sus consecuencias, y hasta para ‘naturalizar’ la filosofía. La teoría de la racionalidad es uno de los aspectos más notables del modelo de científicidad que pueden mostrar las ciencias sociales, y uno de los más próximos a los de las ciencias naturales por su poder explicativo, formalización, cuantificación y grado de contrastación.

La noción de valor económico entendida como utilidad no desempeñaba papel alguno en la teoría clásica –en Smith, Ricardo, Malthus, Marx o J. S. Mill. En estos autores el valor es valor trabajo, una propiedad objetiva y cuantificable de la misma manera –escala- que tiempo o longitud. Hubo, sin embargo, un cambio revolucionario que convirtió el valor trabajo en utilidad. Esta noción en un principio también fue una propiedad cuantificable, pero subjetiva, relativa a cada

sujeto. El concepto de utilidad había aparecido de una manera notoria en el ámbito de la ética. Uno de sus representantes más destacados fue Jeremy Bentham. Propuso un concepto de utilidad objetivo y cuantificable que fue compatible (como teoría ética) con la economía clásica si bien no desempeñó papel alguno en el sistema de conceptos específicamente económicos. Así fue posible que John Stuart Mill, uno de los que condujo a la formulación más acabada del utilitarismo (ético) de su tiempo, como economista desarrolló y expresó sistemáticamente la teoría del valor (trabajo) económico.

Aunque tuvo precedentes, fue William Stanley Jevons quien revolucionó la teoría económica basándola en el concepto de utilidad que tomó de Bentham si bien lo modificó al volver la utilidad relativa al agente, un concepto subjetivo, y lo aplicó a la economía. El potencial de la teoría de la utilidad como una teoría de la acción, tal cual la había formulado Bentham se mostró plenamente en el ámbito económico.

Sin embargo, para formularse cabalmente como teoría de la decisión, debería sufrir otra transformación impulsada por la imposibilidad de deducir de cualquier orden de bienes

económicos una función de utilidad que implicara una escala de medición proporcional igual a la de peso o longitud. La utilidad se *ordinaliza* –implica una escala ordinal- y se vuelve obvio que la estructura de la decisión del agente económico ahora puede extrapolarse a otros campos, a cualquiera, de la acción humana. La dificultad hasta este momento era la imposibilidad de sostener que toda acción moral, política, social, podía cuantificarse con una escala proporcional, es decir, que se pudiera afirmar cuánto mayor es una acción que otra. Este último paso se debe a Vilfredo Pareto.

Este trabajo se propone mostrar brevemente, primero, que la teoría de la decisión es el resultado de una revolución en la teoría económica, no el resultado acumulativo del desarrollo de la teoría económica clásica. Segundo, el desarrollo de la teoría de la utilidad en la ética –como se presenta en Bentham- para volverla aplicable a la economía –como la propuso Jevons-, supone implícitamente la formulación de una teoría de la decisión o de la acción general, no puramente económica. Tercero, la cuantificación ordinal de la noción de utilidad y la prueba de que es suficiente para deducir las consecuencias económicas conocidas, da lugar a

una teoría de la decisión en sentido estricto extrapolable a otras disciplinas.

A continuación se expone en II. la noción de utilidad en J. Bentham que posteriormente influirá en la construcción de la noción moderna de valor económico. En III. la teoría económica clásica del valor y la compatibilidad con la noción de utilidad benthamita. En IV. la teoría marginal del valor y la subjetivización de la utilidad a través de la obra de W. S. Jevons. En V. el último paso necesario para la formulación de una teoría general de la acción –teoría de la decisión–, la ordinalización de la utilidad por V. Pareto. Finalmente en VI. señalaremos algunas conclusiones.

II. Un precedente en la construcción del concepto de valor económico: Jeremy Bentham

J. Bentham, patriarca de los utilitaristas, formó parte de un grupo de tres generaciones: él mismo, James Mill y John Stuart Mill, que influenciaron el pensamiento y la acción política inglesa, entendida esta última de manera amplia que incluye lo que actualmente denominamos política económica. Básicamente eran reformadores no revolucionarios, “they

propose to remedy abuses but do not as yet even contemplate a really revolutionary change.” [Stephen, cap.I, p.1]

Formuló la ley básica del obrar: *actuar de acuerdo al propio interés, que se manifiestan en la búsqueda del placer y en evitar el dolor*. O en sus propias palabras con las que comienza *An Introduction to the Principles of Morals and Legislation*,

Nature has placed mankind under the governance of two sovereign masters, *pain* and *pleasure*. It is for them alone to point out what we ought to do, as well as to determine what we shall do. On the one hand the standard of right and wrong, on the other the chain of causes and effects, are fastened to their throne. They govern us in all we do, in all we say, in all we think: every effort we can make to throw off our subjection, will serve but to demonstrate and confirm it. In words a man may pretend to abjure their empire: but in reality he will remain subject to it all the while. The *principle of utility* recognizes this subjection, and assumes it for the foundation of that system, the object of which is to rear the fabric of felicity by the hands of reason and of law. [cap.1, punto I]

Así, el principio del interés equivale al de la felicidad. Para los utilitaristas lo que mueve al hombre a actuar es la búsqueda de placer. Mediante el principio de utilidad se aprueba o desaprueba una acción según la tendencia a aumentar o disminuir el placer.

Este principio supone que se entiende qué significa utilidad. La propiedad que produce placer se denomina *utilidad*,

By utility is meant that property in any object, whereby it tends to produce benefit, advantage, pleasure, good, or happiness, (all this in the present case comes to the same thing) or (what comes again to the same thing) to prevent the happening of mischief, pain, evil, or unhappiness [Bentham, cap. 1, punto III]

Estas afirmaciones son muy claras respecto del estatus ontológico de la utilidad: es una propiedad objetiva de las cosas. No es una reacción o estado de un sujeto provocado por un objeto; no es relativa al sujeto como lo será para el marginalismo económico. La objetividad de la utilidad constituye la base para que se pueda medir de la misma manera que la longitud, el tiempo o el peso.

El individualismo radical implícito en el utilitarismo es matizado porque reconocen que los individuos poseen sentimientos sociales, especialmente la simpatía que tendrá en A. Smith un representante destacado. Mediante los sentimientos sociales se reconoce que los demás también buscan el placer. De esta manera, el hombre debe obrar para alcanzar el placer para el mayor número de personas, es decir, *la felicidad máxima, obrar de manera de alcanzar la mayor felicidad para el mayor número*, es la medida de lo correcto e incorrecto. Este principio básico,

Bentham himself attributes the authorship of the phrase to Beccaria or Priestley. The general order of thought to which this theory belongs was of course not the property of any special writer or any particular period. Here I need only observe that this embodiment of the general doctrine of utility or morality had been struck out by Hutcheson in the attempt (as his title says) 'to introduce a mathematical calculation on subjects of morality.' This defines the exact reason which made it acceptable to Bentham. For the vague reference to utility which appears in Hume and other writers of his school, he substituted a formula, the terms of which suggest the possibility of an accurate quantitative comparison of different

sums of happiness. In Bentham's mind the difference between this and the more general formula was like the difference between the statement that the planets gravitate towards the sun, and the more precise statement that the law of gravitation varies inversely as the square of the distance. Bentham hoped for no less an achievement than to become the Newton of the moral world. [Stephen, Cp.III,p.2]

Bentham expuso una aritmética de los placeres basada en dos principios: 1. Todos los placeres son cualitativamente iguales, por lo tanto pueden ser medidos; 2. Los placeres de diversas personas pueden ser comparados entre sí.

El placer o el dolor, para un individuo o en sí mismos, pueden medirse teniendo en cuenta los siguientes elementos o dimensiones: la *intensidad*, *duración*, *certeza*, *proximidad* y *seguridad* [Bentham, cap. IV, punto II]. La medición del placer o dolor para un conjunto de personas debe agregar estas dos dimensiones suplementarias: *fecundidad* y *pureza* [Bentham, capIV, punto IV]. Indudablemente está pensando en valores aditivos, en términos contemporáneos, en una escala proporcional,

[Para un conjunto de individuos] *Sum up* the numbers expressive of the degrees of *good* tendency, which the act has, with respect to each individual, in regard to whom the tendency of it is *good* upon the whole: do this again with respect to each individual, in regard to whom the tendency of it is *good* upon the whole: do this again with respect to each individual, in regard to whom the tendency of it is *bad* upon the whole. Take the *balance* which if on the side of *pleasure*, will give the general *good tendency* of the act, with respect to the total number or community of individuals concerned; if on the side of pain, the general *evil tendency*, with respect to the same community. [Bentham, cap.4, punto 6]

Como conclusión de la exposición de su punto de vista resulta interesante preguntarse por el alcance del utilitarismo benthamita. Si una acción produce placer (o más placer que dolor) es buena; este es el punto de vista ético. Ahora bien, si esta acción se selecciona porque maximiza la utilidad sería una acción económica (siempre que los fines perseguidos tengan precios y el agente un presupuesto para gastar); si se maximizan las preferencias sociales sería una acción político-social y así siguiendo. La teoría de la utilidad es una teoría ética pero además puede considerarse una teoría económica,

sociológica, política, etc. Por supuesto, no se espera que Bentham enunciara la teoría económica o política en los términos –muy posteriores– que se acaba de hacer, simplemente se ilustró posibles ampliaciones del utilitarismo del campo de la ética a otros dominios disciplinares.

¿Cuáles de estas teorías formuló Bentham? Sin lugar a dudas una teoría ética, pero sus escritos sugieren, aunque no construyó explícitamente ni se lo propuso, al menos una teoría política o sociológica. Sus referencias a los dos maestros soberanos que la naturaleza impuso al hombre, sugiere que consideraba el utilitarismo una teoría o una filosofía capaz de explicar no solo las conductas éticas sino las conductas políticas, económicas, etc.; una teoría de la acción.

De hecho, su reformismo lo comprometió con posiciones que no se reducían a aspectos exclusivamente éticos. En la práctica propuso reformas socio-políticas detalladas que expresan propuestas enmarcadas en y construidas con elementos de una concepción utilitarista más abarcativa que la ética.

De cualquier modo, los conceptos de utilidad y valor de intercambio todavía no están vinculados como lo fueron a partir de Jevons constituyéndose en la *marca de fábrica* del marginalismo. Las posiciones de Bentham tienen interés para explicar esa conexión, y curiosamente han sido retomadas en la literatura económica contemporánea, visto los problemas empíricos de la teoría de la decisión, para reinterpretar utilidad no simplemente como preferencia sino con contenido en el sentido benthamita [cfr. Kahneman *et al.*, 1997,2005].

Mientras tanto, estos utilitaristas vuelven compatible los principios económicos con la moral. Esta compatibilidad resulta muy clara en Smith mismo, para quien el individuo que piensa solo en sí mismo, en su propio interés, sin embargo providencialmente en este sistema consigue ayudar, usualmente de manera no deliberada, a los demás. La moral utilitarista no se opone a, sino que por el contrario, justifica, los principios económicos, especialmente que cada agente debe atender a sus propios intereses.

III. La teoría clásica del valor económico: el trabajo

Los economistas clásicos intentaron explicar un hecho básico de los fenómenos económicos: ¿cómo se producía el intercambio de bienes? La causa del intercambio son las necesidades o deseos de los agentes y la cualidad que deben tener los bienes para poder ser intercambiados es la igualdad de valor. El valor es la propiedad empírica que explica el intercambio. Pero no bastaba que tuvieran valor sino la misma cuantía, es decir, se intercambian bienes del mismo valor. Lo cual implica poder medirlos y establecer la unidad de valor en el intercambio. La determinación de la propiedad y de su medición son dos problemas distintos. Los economistas clásicos más destacados en la elaboración de la teoría del valor fueron A. Smith, D. Ricardo y K. Marx.

Adam Smith

Realiza una distinción, que todos los clásicos mantuvieron, entre valor de uso y valor de cambio. Valor de uso es la utilidad que representa un bien, valor de cambio la posibilidad de intercambiarlo por otros bienes, "The word VALUE, it is to be observed, has two different meanings, and sometimes express the utility of some particular object, and sometimes the power of purchasing other goods which the possession of

the object conveys.” [Smith, p.44] Esta distinción aparentemente surgió de la observación de la llamada paradoja del valor. Hay bienes que tienen gran valor de uso e inversamente poco o ningún valor de cambio, o por el contrario, otros tienen mucho valor de cambio y poco valor de uso. El agua es un ejemplo del primer caso, un diamante del segundo. Para investigar los principios que rigen el intercambio solamente es interesante investigar el valor de cambio respecto de su naturaleza, la medición y las partes que componen el precio de los bienes [Smith, p.46].

Para Smith el valor de un bien se relaciona con la cantidad de trabajo para producirlo. Se debía considerar dos casos, el primero, es la clase de trabajo que se encuentra en un período específico del desarrollo económico; el segundo, luego que aparece la división sistemática del trabajo. La teoría del valor trabajo en el primer caso es la cantidad del trabajo insumido para obtenerlo y correspondería al estado de la naturaleza de Hobbes, Locke o Rousseau [cfr. Olivera, p.8]. Así Smith afirma [p.65],

In that early and rude state of society which precedes both the accumulation of stock and the appropriation of land, the

proportion between the quantities of labour necessary for acquiring different objects seems to be the only circumstance which can afford any rule for exchanging them for one another.

Aduce a modo de prueba el famoso ejemplo del ciervo y el castor. Así, si en un pueblo de cazadores cazar un castor cuesta el doble de trabajo que cazar un ciervo, el castor será intercambiado por o valdrá dos ciervos [Smith, p.65].

En este estado de la sociedad, el factor clave es que la totalidad del producto pertenece al trabajador “and the quantity of labour commonly employed in acquiring or producing any commodity, is the only circumstance which can regulate the quantity of labour which it ought commonly to purchase, command, or exchange for.” [Smith, p.65]

En el segundo caso, el valor de un bien es, en términos de Olivera [p.7], “la cantidad de trabajo que puede economizar a su poseedor mediante el cambio.” El valor de un bien que alguien posee para intercambiarlo por otro,

is equal to the quantity of labour which it enables him to purchase or command. Labour, therefore, is the real measure

of exchangeable value of all commodities. The real price of every thing, what every thing really costs to the man who wants to acquire it, is the toil and trouble of acquiring it. [Smith, p.47]

En este estado de cosas el producto total del trabajo no pertenece al trabajador quien debe compartirlo con el dueño del capital que lo emplea. De una manera más precisa, en una economía capitalista el salario es menor que el producto total del trabajo, es decir, el trabajo economizado por la venta del producto es mayor que la cantidad de trabajo insumida por su producción [cfr. Smith, pp.67-8 y Olivera, pp.9-11].

David Ricardo

La teoría del valor desarrollada por Ricardo es más compleja pero más simple y precisa en su exposición y en lo fundamental continúa el tratamiento de Smith; el valor es considerado como una propiedad objetiva, una cantidad de trabajo.

Ricardo comienza el capítulo I de sus *Principles* citando la distinción de Smith entre *valor de uso* y *valor de cambio* y explicando “la paradoja del valor”. Inmediatamente relaciona

ambos, afirmando que si un bien no fuera útil, es decir, si no tuviera valor de uso tampoco tendría valor de cambio. A continuación expresa que el valor de cambio de los bienes deriva de dos fuentes: "from their scarcity, and from the quantity of labour required to obtain them." [Ricardo, p.5]. Hay bienes cuyo valor está determinado por su escasez, por ejemplo, las obras de arte. Otros bienes pueden ser producidos sin límite y la competencia actúa sobre ellos sin restricciones. A estos últimos se referirá Ricardo.

Para éste, el principio general del valor "depends almost exclusively on the comparative quantity of labour expended" [Ricardo, p.6]. Como señala Olivera "es la generalización de la tesis especial de A. Smith: la determinación del valor de cambio por el trabajo insumido"[Olivera, p.12]. Luego de una extensa cita de Smith en la que explica este principio en los grupos en estado de naturaleza Ricardo afirma,

That is really the foundation of the exchangeable value of all things, excepting those which cannot be increased by human industry, is a doctrine of the human industry, is a doctrine of the utmost importance in political economy; for from no source do so many errors, and so much difference of opinion in that

science proceed, as from the vague ideas which are attached to the word value.

If the quantity of labour realized in commodities regulates their exchangeable value, every increase of the quantity of labour must augment the value of that commodity on which it is exercised, as every diminution must lower it. [Ricardo, p.7]

Aunque su concepción presenta dificultades, distingue claramente entre diferentes aspectos: las causas de que los bienes tengan valor, la propiedad empírica que denomina valor, la regla de intercambio de los bienes, y la unidad para medir el valor. El primer aspecto reside en la vinculación entre valor de uso y valor de cambio, y acabamos de definir el aspecto empírico, la propiedad, que denomina valor, el trabajo insumido en la obtención de un bien reproducible sin límite. La regla de intercambio de los bienes no escasos es simple: se intercambian según igualdad de valor. Recurriendo al conocido ejemplo de A. Smith, si en una comunidad de cazadores atrapar un castor supone el doble de trabajo que atrapar un ciervo, un castor se intercambiará por dos ciervos. La formulación de la regla de intercambio supone conocer la propiedad empírica que se compara, es decir, que se mide.

La caracterización de la noción valor-trabajo presenta algunas complicaciones. En las primeras etapas de la sociedad aparece clara, de la manera que la mostraba el ejemplo del valor del ciervo y del castor. El valor de un bien se debe al trabajo directo aplicado a su producción y se puede incrementar por el empleo de útiles. En el ejemplo de la caza, el arco y la flecha, las trampas. O en el caso de la producción moderna, las máquinas mediante las cuales se producen los bienes. En este caso,

still the same principle would hold true, that the exchangeable value of bestowed on their production; not on their immediate production only, but on all those implements or machines required to give effect to the particular labour to which they are applied. [Ricardo, p.14].

Otra determinación del trabajo y fuente de variaciones del valor es la calidad del trabajo y el aprendizaje para realizarlo [Ricardo, pp.11-12]. La cantidad de trabajo se basa, además, en la destreza o capacitación. Cita en su apoyo una observación en el mismo sentido realizada por A. Smith. En fin, a medida que la sociedad se desarrolla la producción de un bien insume más trabajo de formas muy diversas y

complejas pero el principio de su valor es el mismo, considerar el trabajo total insumido, es decir, el trabajo directo más el trabajo acumulado utilizado para producirlo (transporte de materiales, máquinas, capacitación, etc.)

Pero el trabajo acumulado mencionado simplemente se puede sumar? Aquí hay un problema de agregación que Ricardo consideró. Para Olivera una de las mayores contribuciones de la teoría ricardiana del valor-trabajo es "la inclusión del tiempo. Ricardo reconoció de un modo explícito la influencia del tiempo en la determinación del valor y, aun más, la vinculó a la tasa de interés" [Olivera, p.12].

Al comienzo de la sección II de la Parte Primera [p.11] Ricardo expresa muy claramente que el trabajo es el fundamento del valor y que "la cantidad relativa de trabajo" determina el valor relativo de los bienes. En palabras contemporáneas, "la cantidad relativa de trabajo" posibilita comparar, establecer un orden, es decir, un sistema comparativo entre los bienes sobre la base de la relación empírica objetiva, 'tener más valor-trabajo que'. Ricardo no confunde esta estimación con otro problema, el de la determinación de cuánto más trabajo hay en el bien x que en el bien y. La

obtención de una escala proporcional significa determinar la unidad de medición. Ricardo [sección VI, p.27 y ss.] argumenta acerca de la imposibilidad de obtenerla. La dificultad reside en encontrar una unidad invariable.

El problema de la teoría del valor-trabajo de Ricardo es básicamente la misma de todos los clásicos, incluido Marx: hacer comparables distintas formas de trabajo. No hemos señalado exhaustivamente las formas que determinan el valor de los bienes no escasos, pero a las pocas mencionadas las consideramos suficientes para señalar la naturaleza del problema. Su formulación del valor-trabajo, a pesar de las dificultades, es la más precisa entre los individualistas clásicos.

El predicado valor-trabajo en Ricardo, al igual que en el resto de los clásicos, es individualista porque las formas de trabajo son producidas por cada uno de los agentes económicos. No le presta atención especial, como lo hace A. Smith, o aun más claramente, como J. Stuart Mill. En cambio, dedica entero el capítulo 1 de sus *Principios* a la definición del valor mediante predicados objetivos: trabajo directo incorporado a

un bien; trabajo utilizado para producir un útil o máquina; trabajo empleado para capacitar a un trabajador, etc.

John Stuart Mill

No es la intención en esta sección examinar con0 detalle la teoría del valor de J. Stuart Mill, que pertenece a la tradición del valor-trabajo, aunque posea diferencias con los autores precedentes. Interesa solo examinar aquellos conceptos por los cuales se constata que no se había dado –o no era posible- la conexión entre el utilitarismo de Bentham y el valor (de cambio) de un bien.

En sus *Principios de Economía Política* J. S. Mill comienza la parte dedicada a los intercambios, proponiendo denominar a la Economía Política con los términos *Cataláctica*, o ciencia de los intercambios, o también *Ciencia de los Valores*. Al examinar la distinción entre valor de uso y valor de cambio que expusimos en Smith, acuerda que hay bienes que tienen el mayor valor de uso y muy poco o ningún valor de cambio pues empleamos muy poco o ningún trabajo para conseguirlos, el agua de lluvia o el aire. En cambio, disiente que haya bienes que puedan tener gran o el mayor valor de

cambio y poco o ningún valor de uso, como un diamante. En economía el valor de uso es la satisfacción de un deseo o la realización de un propósito; entonces, el valor límite del valor de intercambio es el valor de uso. Nunca puede el valor de cambio ser mayor que el valor de uso. La afirmación de Smith se debe a que utiliza valor de uso en una doble acepción,

This is employing the word use, not in the sense in which political economy is concerned with it, but in that other sense in which use *is opposed to pleasure*. Political economy has nothing to do with the comparative estimation of different uses in the judgment of a philosopher or of a moralist. [Mill, 1909; parte III, cap.1, párrafo 3; cursivas nuestras]

Es claro en este pasaje la separación entre utilidad como es entendida posteriormente por los marginalistas, compatible con la utilidad en el sentido de Bentham, y el valor trabajo. Mill conocía perfectamente el utilitarismo benthamita y pretende desarrollarlo, siendo reconocido como uno de los padres de esta corriente [cfr. Mill, 2002]. Difería principalmente de Bentham porque reconocía diferentes utilidades, es decir, había diferentes clases de utilidades y porque no se podían medir.

La noción de trabajo sufrió un cambio profundo, en parte debido a las dificultades de medir cantidades de trabajo diferentes. Con la aparición de las teorías neoclásicas se produjo un completo desplazamiento ontológico del valor (económico); desde entonces se lo concibe como preferencia.

IV. La teoría marginal del valor económico: la utilidad

El libro *Principles of Economy* de *W. S. Jevons*, publicado en [1871] es el emblema de una revolución que comenzaba en la economía conjuntamente con los *Grundsätze* de Menger [1871] y, tres años más tarde, los *Éléments d'économie politique pure* de Walras [1874].

Para que no quede duda de su posición original, citamos el Prefacio de su obra mencionada,

Repeated reflection and inquiry have led me to the somewhat novel opinion, that *value depends entirely upon utility*. Prevailing opinions make labour rather than utility the origin of value; and there are even those who distinctly assert, that labour is the *cause* of value. I show, on the contrary, that we have only to trace out carefully the natural laws of the variation of utility, as depending upon the quantity of

commodity in our possession, in order to arrive at a satisfactory theory of exchange, of which the ordinary laws supply and demand are a necessary consequence. [Jevons, cap. 1, p.2]

Jevons es plenamente consciente de la novedad de su teoría ya en el primer bosquejo presentado en su conferencia en la British Association [1863]. En su época predominaba el punto de vista que la economía estaba bien fundamentada y sistematizada y algunos de sus capítulos, como el del valor, no había nada que se pudiera clarificar –tal como lo había afirmado J.S.Mill. Emplea mecanismos persuasivos muy específicos para que la audiencia acepte los nuevos puntos de vista; recurre especialmente a la concordancia de sus nuevas teorías con los hechos, a los precedentes de esas concepciones que se encuentran en autores anteriores y en la nueva manera de presentar de modo más preciso los conceptos económicos mediante el lenguaje matemático como las ciencias desarrolladas.

Jevons afirma que intenta construir la economía como un cálculo del placer y el dolor (1871, p.vii; 1866, punto 2). Estos dos son en última instancia los únicos motivos de las

acciones de los seres humanos. Los motivos que estudia la economía son los que guían las acciones ligadas a la satisfacción de las necesidades humanas; de otra manera, no todo placer y dolor es objeto de la economía, quedan excluidos de su ámbito los placeres (dolores) morales, estéticos, etc. Sigue el principio de utilidad que ya había formulado Bentham, maximizando la felicidad al mínimo costo, "The theory which follows is enterely based on a calculus of pleasure and pain; and the object of Economy is to maximise happiness by purchasing pleasure, as if were, at the lowest cost of pain." [Jevons, 1871, p.27].

Considera los sentimientos de placer y dolor como magnitudes de acuerdo a los aspectos que distinguió Bentham al cual cita directamente (intensidad, duración, certeza y proximidad), aunque es mucho más preciso y cuidadoso al establecer que no existe una unidad de medida para ellos –lo cual no impide que la economía sea una ciencia matematizable [Jevons, 1871, p.11]. Si bien, entonces, los sentimientos no se pueden medir directamente, sin embargo, se pueden estimar comparativamente sus respectivas cantidades por sus efectos cuantitativos en las cosas [Jevons, 1871, p. 14-5]. Se puede estimar la igualdad o

diferencia en la cantidad de placer o dolor por las variaciones en las decisiones en la mente humana; o sea, por las decisiones o deseos determinamos el placer o el dolor y sus oscilaciones en cantidad quedan registradas en los precios de mercado [Jevons, 1871, p.15]. Si comprendemos cabalmente a Jevons, su afirmación es análoga a lo que hoy se afirma comúnmente de las preferencias, que no se pueden conocer directamente y son estimadas indirectamente por sus efectos, por las elecciones que realiza un agente.

Gran parte del trabajo empleado por una comunidad consiste en la producción de alimentos, ropas, construcciones para satisfacer las necesidades humanas y son estos objetos las causas del placer y dolor. Por esta razón y por lo señalado precedentemente, es conveniente transferir nuestra atención rápidamente a estos objetos que son la fuente de placer y dolor.

En este momento es conveniente introducir una precisión terminológica realizada por Jevons. Por *bien económico* ('commodity') se entiende cualquier objeto, sustancia, acción o servicio que produzca placer o evite el dolor. Por *utilidad* se denota la propiedad abstracta por la cual algo es un bien

económico [Jevons, 1871, cap.III] y señala que otorga a este término el mismo significado que Bentham, solamente que le interesa restringirlo a su uso en economía.

De esta manera, la economía debe investigar y basarse en las condiciones de utilidad, por lo cual hay que examinar las necesidades y deseos de los individuos. Esta afirmación significa lo mismo que decir que la economía debe comenzar elaborando una teoría del consumidor –contra la opinión de los clásicos que expusiera explícitamente J. S. Mill, citado por Jevons [1871, cap.III], "Political economy, has nothing to do with the consumption of wealth, further than as the consideration of it is inseparable from that of production, or from that of distribution."

Una de las características más notables de la utilidad es que no consiste en una propiedad intrínseca de los objetos, "It is better described as *a circumstance of things* arising out of their relation to man's requirements." [Jevons, 1871, cap.III]. Así, el diamante todavía no descubierto no posee valor para ningún individuo. Actualmente podríamos caracterizarla más precisamente afirmando que es una propiedad relativa a un individuo, una propiedad subjetiva. Es casi innecesario

enfatar la inversión copernicana que ha realizado Jevons, el valor de un bien económico no queda determinado por el trabajo sino por su utilidad, "Repeated reflection and inquiry have led me to the somewhat novel opinion, that *value depends entirely upon utility*. Prevailing opinions make labour rather than utility the origin of value" [Jevons, 1871, p.2].

Otra característica notable de los bienes económicos es la variación de su utilidad. Cierta cantidad de agua es indispensable para la vida; cantidades adicionales poseen varios grados de utilidad; a partir de cierta cantidad la utilidad tiende a cero; y más cantidades de agua pueden causar desutilidad. Es intuitivo que la utilidad de un bien varía de acuerdo a la cantidad que ya se posee de ese bien. Esto le conduce a formular la diferencia usual en nuestros días entre la utilidad total y la utilidad marginal (Jevons la denominaba *grado de utilidad*).

Con los conceptos introducidos hasta este momento está capacitado para abordar el intercambio. Es uno de los procesos más importantes para maximizar la utilidad y ocupaba un lugar central en la economía clásica aunque era tratado siempre en una etapa posterior al trabajo.

Jevons por razones que se han explicitado antes afirma que, “I find it both possible and desirable to consider this subject [el intercambio] before introducing any notions concerning labour or the production of commodities.” [Jevons, 1871, cap. IV]. En la economía clásica el trabajo era la medida del valor y la igualdad de valor la explicación del intercambio. El valor era cantidad de trabajo, una propiedad objetiva, como Mill claramente precisó, “The value of a thing means the quantity of *some other thing*, or of things in general, which it exchanges for.” [Mill, Libro III, cap. 4; subrayado nuestro], es una cosa. Jevons estará totalmente en desacuerdo pues para él es una propiedad relativa, no intrínseca. Vale la pena citarlo al respecto,

Now, if there is any fact certain about exchange value, it is, that it means not an object at all, but a circumstance of an object. Value implies, in fact, a relation; but if so, it cannot possibly be *some other thing*. A student of Economics has no hope of ever being clear and correct in his ideas of the science if he thinks of value as at all a *thing* or an *object*, or even as anything which lies in a thing or object. Persons are thus led to speak of such a nonentity as *intrinsic value*. There are, doubtless, qualities inherent in such a substance as gold

or iron which influence its value; but the word Value, so far as it can be correctly used, merely expresses *the circumstance of its exchanging in a certain ratio for some other substance*. [1871, cap.iv]

Claramente si el valor es una ratio no puede considerarse cosa alguna, ni algo de un bien. ¿Cuál es la ratio (el valor) de una tonelada de trigo?, no tiene significado.

Recordemos la distinción que realizaba Smith respecto al valor de uso y valor de cambio, para Jevons solo tiene sentido si se precisan estos significados tomados del uso común y se entiende que valor de uso es la utilidad total y el valor de cambio la ratio de intercambio.

V. Del valor económico a una teoría general de la acción: la teoría de la decisión

La teoría económica tuvo un gran desarrollo debido a los avances de Jevons, pero también de manera independiente por los de Walras y los de Menger. Ellos sentaron las bases e influenciaron trabajos que dieron lugar a una tradición

específica de resolución de problemas en el ámbito de la economía, fueron los fundadores del marginalismo y de allí se originaron subescuelas como los austriacos. Se explicaron los fenómenos económicos tradicionales en los términos marginalistas como la teoría de la producción, el trabajo, el dinero, y se propusieron conceptos y teorías novedosas como la teoría de la competencia perfecta, el duopolio, y por citar un concepto que nos interesa, las curvas de indiferencia introducidas por Edgeworth para explicar la demanda individual. Todos suponían la utilidad cardinal [cfr. en Rebeyrol, pp.39-47 el caso de Walras], es decir, la medición mediante una escala proporcional de la de la utilidad –o al menos posibilidad teórica de esa medición. La contribución decisiva de V. Pareto consistió en mostrar la imposibilidad en ciertos casos de poder utilizar la utilidad cardinal para resolver problemas y, en cambio, mostrar la posibilidad de deducir la economía conocida a partir de la noción de preferencia/indiferencia y obtener una función de utilidad ordinal.

Los dos primeros capítulos de su *Manual* [1927] tienen que ver directamente o de manera relacionada con cuestiones metodológicas. Se exponen algunas no por novedosas sino

para dar un contexto a las posiciones que nos interesan remarcar. Comienza meticulosamente caracterizando la ciencia e incluyendo la economía dentro de esta clase de conocimiento y a su vez dentro de las ciencias sociales. La economía política es una ciencia social y como toda ciencia se caracteriza por formular leyes en su dominio,

Les actions humaines présentent certaines uniformités, et c'est seulement grâce à cette propriété qu'elles peuvent faire l'objet d'une étude scientifique. Ces uniformités portent encore un autre nom ; on les appelle des *lois*. [Pareto, p.5]

Las leyes tienen una estructura condicional, es decir, son verdaderas bajo ciertas condiciones, "Il n'y a pas de proposition qu'on ne puisse certifier vrai sous certaines conditions" [Pareto, p.9].

De la afirmación incuestionable que nunca se puede conocer completamente un fenómeno con sus infinitas propiedades extrae la conclusión que las teorías siempre tienen carácter aproximativo, es decir, estudian modelos idealizados que nunca se puede decir cuánto se alejan de la realidad. De esta manera una teoría puede ser útil para alcanzar cierto fin y

otra para uno distinto. Sin embargo, las teorías se descartan cuando no concuerdan con la realidad, “elles doivent être d'accord avec les faits” [Pareto, p.11]. El ajuste con la realidad es la piedra de toque de las teorías que vuelve a la ciencia un “perpetuo devenir” al obligar a rechazar las teorías falsas.

El ajuste se puede dar de dos maneras, como las ciencias semejantes a la física o a la química que pueden recurrir a la experiencia (es decir, al control experimental) o de la manera que se da en la economía política, la astronomía, que recurren a la observación [Pareto, p.16]. A pesar de esta diferencia ambas clases de ciencias tienen en común la búsqueda de leyes verdaderas, o sea, que concuerden con la realidad valiéndose de la experiencia.

Las ciencias sociales estudian aspectos objetivos así como aspectos subjetivos, por ejemplo, los gustos de los hombres. La economía política incluye el estudio de ambos aspectos y su objeto es el estudio del equilibrio económico considerando las sensaciones (reacciones) de un individuo frente a condiciones diferentes y determinar cuál de esas condiciones elegirá guiado exclusivamente por su interés personal

[Pareto, p.148-49]. Con este fin se deben estudiar los gustos, los obstáculos para dar satisfacción a estos gustos y la manera que a partir de ambos el individuo obtiene el equilibrio.

Pareto reseña brevemente cómo se sometieron a cálculo los gustos, deduciéndolos del placer que determinadas cosas producen a los individuos, denominado *valor de uso* o *utilidad*. Esta aproximación tuvo dos defectos, se la consideró una propiedad objetiva y no se advirtió que dependía de las cantidades consumidas. Jevons rectifica estas teorías y utilidad toma un sentido técnico alejado del significado común. Esta es la razón por la cual Pareto prefiere denominarla *ofelinidad* [p.157]. Con la noción de función de utilidad se construía la utilidad total, la marginal, el equilibrio del consumidor, las condiciones de intercambio, y así siguiendo. Con las curvas de indiferencia –noción introducida por F.Y. Edgeworth- también se construían las nociones usuales en economía que conducían al equilibrio del consumidor, al intercambio, etc. La noción de utilidad supuesta por Jevons, Walras, Edgeworth y los restantes hasta Pareto fue la utilidad cardinal, es decir, cuantificada como una función lineal o en términos más modernos,

suponía una escala proporcional. Ahora bien, nunca fueron demostrados los respectivos teoremas de representación y unicidad. El primer obstáculo fue que nunca se encontró cuál era la unidad empírica de medición. Pareto en el *Apéndice* de su obra [ver Anexo I] demuestra, por el contrario, que en general siempre que el consumo de un bien x, depende simultáneamente del consumo de otros bienes y, z, ... esta función de utilidad es indeterminada!,

La théorie de la ophélimité a reçu un nouveau perfectionnement. (...) Nous avons admis que cette chose appelée *plaisir, valeur d'usage, utilité économique, ophélimité*, était une quantité ; mais la démonstration n'en a pas été donné. Supposons cette démonstration faite, comment ferait-on pour mesurer cette quantité ? C'est une erreur de croire que d'une façon générale, on puisse déduire de la loi de l'offre et la demande la valeur de l'ophélimité. (...) quand la ophélimité de une merchandise depend (...) [de la consommation des autres] la ophélimité reste indéterminée, même après qu'on a fixé l'unité qui sert à la mesurer. [Pareto, p.159].

Hicks [1934] afirma que en la teoría del valor había sido desarrollada intensamente en la generación de Jevons y Marshall pero que desde 1900 no había habido ningún avance significativo hasta la aparición del *Manuel* de Pareto que expone la más completa teoría estática del valor. Su aporte más significativo consistía en “his demonstration of the incommensurability of utility” [p.52] y la reemplaza por una escala de preferencias que conduce a un índice de utilidad pero no nos habilita a obtener a partir de la escala de preferencias una función de utilidad cardinal. Se puede pasar de una función de utilidad (cardinal) a una escala de preferencias pero la inversa no está garantizada [cfr. Hicks y Allen, p.52-3].

La propuesta de Pareto expresada en términos actuales afirma que para la teoría económica lo que interesa son las curvas de indiferencia, o mapa de indiferencia, y se puede prescindir sin más de la superficie de indiferencia completa y retener de esta solo las curvas de indiferencia [cfr. Pareto, pp.170-1 que expresa lo mismo mediante la imagen de la colina del placer].

¿Este cambio fue revolucionario como el paso del valor trabajo a la utilidad? La respuesta que aquí se sostiene es

negativa: no consistió en un cambio revolucionario pues se obtienen los mismos resultados. Es cierto que la utilidad como cualidad empírica desaparece, sea objetiva o subjetiva, a favor de la relación empírica 'ser preferido', lo cual significa un cambio ontológico. Pero este cambio implica modificaciones menores, implica reescribir algunas nociones de la teoría, pero no desaparecen resultados o se incorporan nuevos. Un cambio significativo, y de este estilo son los restantes, es el paso de la utilidad marginal decreciente a la tasa marginal de sustitución. La ordinalización de la utilidad fue muy importante pero no tiene el alcance de una revolución científica.

L. Robbins lo expresa de manera muy compacta y clara; vale la pena transcribir su exposición,

No se requiere un gran conocimiento del análisis económico moderno para comprender que la teoría del valor descansa en el supuesto de que las diferentes cosas que un individuo desea tener poseen para él una importancia diversa y pueden ser dispuestas, por consiguiente, en un orden determinado. Esta noción puede expresarse en varias formas y con diversos grados de precisión, desde el simple sistema de

necesidades de Menger y de los primeros austriacos, hasta las más refinadas escalas de valoraciones relativas de Wicksteed y Schönfeld y los sistemas de indiferencia de Pareto, Hicks y Allen. Pero, en último análisis, se reduce a que podemos juzgar si diferentes experiencias posibles son para nosotros de una importancia mayor, menor o equivalente. De este hecho elemental de la experiencia podemos derivar la idea de la sustitutibilidad de distintos bienes; de la demanda de un bien en función de otro; de la distribución equilibrada de bienes entre usos diversos; del equilibrio de cambio y de la formación de los precios. Al pasar de la descripción de la conducta de un solo individuo al examen de los mercados, hacemos, naturalmente, otros supuestos subsidiarios: se trata de dos o más individuos; la oferta está en manos de un monopolio o de vendedores múltiples; las personas que intervienen en el mercado conocen o ignoran lo que está sucediendo en otros sectores del mismo; el marco jurídico del mercado prohíbe este o aquel modo de adquisición o de cambio, etc. Suponemos, asimismo, una determinada distribución inicial de la propiedad. Pero el supuesto principal implícito es siempre el de las escalas de valoración de los distintos sujetos económicos. [Robbins, cap.4, párrafo 2]

Una pregunta interesante es ¿por qué razón Pareto sustituyó utilidad por preferencias? No puede deberse al hecho que la utilidad no es medible con una escala proporcional pues no desaparecen los conceptos no cuantificables mediante esas escalas. La razón está en otra parte. Indudablemente la utilidad tenía un largo recorrido, desde Bentham al menos, en que un factor para el lugar central que ocupaba se debía a la persuasión que podía ser medible, si ahora se demostraba que no era el caso en general y se tenía a mano un concepto que desempeñaba las mismas funciones y, además, era menos abstracto -incluso Pareto lo igualaba a lo observable-, usando este nuevo concepto se aseguraba una base empírica más sólida, ligada al comportamiento observable de los individuos. También desempeña un papel el hecho de que la noción básica de la que partía la economía al reemplazar la utilidad parecía ser más simple e indisputable: simplemente un sujeto ordena de acuerdo a sus preferencias un conjunto de bienes. La utilidad es una noción mucho más abstracta, aún cuando se asocie al placer. Un párrafo de Pareto avala esta interpretación,

Grâce a l'usage de la mathématiques, toute cette théorie, telle que nous la développons dans l'Appendice, *ne repose plus*

que sur un fait d'expérience, c'est-à-dire sur la détermination des quantités de biens que constituent des combinaisons indifférentes pour l'individu. La théorie de la science économique acquiert ainsi la rigueur de la mécanique rationnelle ; elle déduit ses résultats de l'expérience, sans faire intervenir aucune entité métaphysique. [Pareto, p.160 ; subrayado nuestro].

Aunque irónicamente poco después también se volvió sospechosa por su grado de abstracción y los ultraempiristas intentaron eliminarla derivándola de la noción indisputablemente segura, observacional, de elección. Este movimiento dio lugar a la denominada Teoría de la Preferencia Revelada [cfr. Samuelson].

A la utilidad le quedó finalmente un papel honorífico en el mainstream, designando la función que vincula las preferencias con la estructura numérica: si $x P y$ entonces $f(x) > f(y)$, P es la relación de preferencia y f es la función de utilidad.

La ordinalización de la utilidad ha tenido otras implicaciones significativas, un resultado inesperado es la aplicabilidad, potencialmente al menos, a otros campos de fenómenos,

The methodological implications of this “ordinal” conception of utility have been discussed elsewhere; they are very far-reaching indeed. By transforming the subjective theory of value into a general logic of choice, *they extend its applicability over wide fields of human conduct*. [Hicks y Allen, p.54; subrayado nuestro]

Las transformaciones de la utilidad condujeron a una verdadera teoría de la elección o de la decisión que no es sino una teoría de la acción humana. Desde entonces la teoría de la decisión económica fue un modelo explicativo exportable a otros dominios de las disciplinas sociales de la racionalidad humana.¹

¹ Paradójicamente, el problema de la cardinalidad vuelve a reaparecer cuando von Neuman y Morgenstern [1967] demuestran que es una condición necesaria de la teoría de la decisión bajo riesgo considerar la utilidad cardinalmente. En este caso solo puede haber medición si además del orden se pueden medir las distancias relativas de los intervalos de preferencia, o sea, asegurar que se puede establecer

VI. Conclusiones

Ha quedado claro en el examen realizado, principalmente histórico, que la noción de utilidad no desempeñó un papel fundamental equivalente a un concepto primitivo, en la economía clásica. Para explicar el fenómeno económico básico del intercambio los economistas clásicos recurrían al valor trabajo que poseía un bien, su igualdad permitía el intercambio. Dificultades internas con la agregación de muy diferentes tipos de trabajos condujo, entre otros muchos factores, a la propuesta de otro concepto para la misma explicación: la utilidad que poseen los bienes. Este concepto estaba disponible desde mucho antes pero aplicado especialmente a fenómenos éticos o políticos, no a la economía. Fue especialmente Jevons quien lo empleó con modificaciones; principalmente lo volvió un concepto subjetivo aunque mantuvo la pretensión de su medición con una escala proporcional. Los cambios que introdujo supusieron una

una escala de intervalos (se puede transformar linealmente) y es una escala más fuerte que la ordinal pero menos que la proporcional.

reformulación de la teoría económica que implicó una revolución: el marginalismo. Esta corriente modifica la ontología mediante, por ejemplo, el valor subjetivo y cambia los conceptos básicos económicos: la manera de entender el valor, el papel primordial de la teoría del consumo antes secundaria o considerada fuera de las explicaciones económicas, las diferentes explicaciones de fenómenos básicos como el intercambio que no se basa justamente en la igualdad de valor (utilidad). La utilidad así concebida pone de manifiesto que los fenómenos económicos se generan a partir de los deseos de los consumidores, y la modelización de las decisiones de consumo mediante la teoría del consumidor inicia el desarrollo de la totalidad de la economía. Esta manera explícita o implícita la necesidad de comenzar con una teoría de la decisión económica del individuo. La imposibilidad para probar en general la medición de la utilidad con una escala proporcional condujo a Pareto a proponer medirla con una escala ordinal. El interés de este paso es fundamental para poder generalizar el modelo económico de decisión de los individuos a otras disciplinas: la psicología, la biología, la politicología, etc. La teoría de la decisión se convierte en una teoría de la racionalidad, es decir, en una teoría general explicativa de la acción de cualquier individuo.

BIBLIOGRAFÍA

Bentham, J. (1823), *An Introduction to the Principles of Morals and Legislation*. Oxford: Clarendon Press, 1907. Reproducida on-line en: <http://oll.libertyfund.org>

Dobb, M. (1975), *Teoría del valor y de la distribución desde Adam Smith –Ideología y teoría económica*. Siglo XXI.

Hicks, J.R.and R.G.D. Allen (1934), "A Reconsideration of the Theory of Value". Economica, New Series,v.1, nº1, February, pp.52-76.

Jevons, W. S.(1866), "Brief Account of a General Mathematical Theory of Political Economy". Journal of the Royal Statistical Society, June, pp.282-87.

_____(1871), *The Theory of Political Economy* . London: Macmillan and Co. Disponible en <http://www.econlib.org/library/>

Kahneman, D., P. Wakker and R. Sarin (1997), "Back to Bentham? Explorations of Experienced Utility". Quarterly Journal of Economics, 112, pp.375-405.

Kahneman, D. and R. Sugden (2005), "Experienced Utility as a Standard of Policy Evaluation". Environmental & Resource Economics, 32, pp.161-181.

Menger, C. (2007) [1871], *Principles of Economy*. L. von Mises Institute. Traducido por J. Dingwalland y B. F. Hoselitz.

Mill, J. S. (1909), *Principles of Political Economy with some of their Applications to Social Philosophy*. London; Longmans, Green and Co. Reproducida on-line en: <http://oll.libertyfund.org>

_____(2002), *El utilitarismo*. Alianza editorial.

Olivera, J.H.G. (1957), *Valor y Trabajo –Estudio sobre la teoría clásica y marxista del valor*. Facultad de ciencias Económicas-UBA (mimeografiado).

Pareto, V. (1927), *Manuel d'economie politique*. M. Giard, PARIS.

Ricardo, D. (1962), *The Principles of Political Economy and Taxation*. J.M.Dent & sons; última reimpression.

Rebeyrol, A. (1999), *La pensée économique de Walras*.Dunod.

Robbins, L. (1944), *Ensayo sobre la naturaleza y la significación de la ciencia económica* . Fondo de Cultura Económica.

Samuelson, P. A. (1938a), "A note on the Pure Theory of Consumer's Behaviour". Economica, February, pp.61-71.

Smith, A. (1981), *An Inquiry into the Nature and Causes and the Wealth of Nations*. Liberty Fund; esta es una reproducción fotográfica de la edición publicada por Oxford University Press de 1976 cuyos editores generales fueron R. H. Campbell y A. S. Skinner.

Stephen, L., *The English Utilitarians- Jeremy Bentham*. Vol.1
En: <http://www.efm.bris.ac.uk/het/bentham/stephen1.htm>

Stigler, G.J. (1950), "The Development of Utility Theory. I". The Journal of Political Economy, v.58, n.4, august; pp.307-27.

_____(1950), "The Development of Utility Theory. II". The Journal of Political Economy, v.58, n.5, october; pp.373-96.

Viner, J.(1925), "The Utility Concept in Value Theory and its Critics". The Journal of Political Economy, v.33, n.4, August, pp.369-37.

_____(1949), "Bentham and Mill: The Utilitarian Background". The American Economic Review, v.39, n.2, March, pp.360-82.

Walras, L. (1988) [1874], *Éléments d'économie politique pure, ou théorie de la richesse sociale*. V. 8, [Oeuvres économiques complètes](#), Economica.

Von Neumann, J., O. Morgenstern (1967) [1944], *The Theory of Games and Economic Behavior*. Wiley.

LA UTILIDAD COMO CUALIDAD HEDÓNICA: SUS AVANCES, LIMITACIONES, Y PRINCIPALES CAMPOS DE INVESTIGACIÓN APLICADA.

Leonardo Ivarola

Bentham postuló que las acciones de los individuos eran guiadas por sensaciones de placer y dolor. A partir de esta premisa consideró que la utilidad de todo bien se hallaba ligada a una cualidad hedónica. Además, asumía la existencia de un legislador que aplique políticas económicas destinadas a generar el máximo placer para toda la comunidad, haciendo mínimo el dolor. La utilidad en sentido hedonista fue dejada de lado posteriormente, principalmente a partir de la tesis de Pareto: solo era relevante conocer la relación de preferencia entre dos o más bienes, y no la magnitud en que uno se prefería sobre otro. Empero, el rol del placer-dolor fue restituido por la corriente de la *economía conductual*, donde se realizaron numerosas investigaciones en torno a lo que estos han denominado *experienced utility* (o utilidad en el sentido de Bentham), para distinguirlo de *decision utility*: el peso de un resultado en una decisión. Tanto la medición como el uso de *experienced utility* han abierto

varias líneas de investigación empíricas, tales como testear la hipótesis de maximización, medir el bienestar subjetivo de la población en general o encontrar un estándar de evaluación de políticas.

Propedéutica.

La utilidad entendida como medida de procesos hedónicos fue propuesta inicialmente por Bentham hacia fines del siglo XVIII, dando origen a la corriente del *utilitarismo*. En esta, se consideraba que los actos de los individuos estaban guiados en torno a dos variables antagónicas: placer y dolor, y dichos actos pretendían buscar maximizar el placer, minimizando el dolor. La misma idea se aplicaba al ámbito de las políticas económicas, donde el legislador aplicaba dichas políticas a los efectos de buscar el máximo bienestar o placer de la comunidad total, es decir, buscaba maximizar la utilidad social.

Bentham propuso una forma de medir el placer y dolor de cada uno de los individuos, teniendo en cuenta todas aquellas variables que generaban estos estados hedónicos. Empero, dicho método no lograba determinar las magnitudes

cuantitativas de placer y dolor, es decir, la cantidad exacta (o aproximada) de placer-dolor que implicaba un acto cualquiera. A grandes rasgos, podríamos ver que en Bentham, todas las decisiones de los individuos *valían por igual*, prescindiendo de esta manera de las diferentes magnitudes de placer-dolor de cada uno de los individuos.

A comienzos de siglo XX, la concepción hedonista de la utilidad fue dejada de lado, principalmente mediante los aportes de Pareto, quien por un lado propuso una definición de utilidad mucho más abstracta, y por otro postuló que en el análisis económico lo relevante eran las relaciones de preferencia, y no en qué medida se prefería un acto a otro. Por ejemplo, dados dos bienes K y Z, la concepción utilitarista asumía que los mismos generaban en el individuo una cantidad de placer, y que esta cantidad era mensurable, de forma tal que la utilidad de K podía ser 8 y la de Z ser 4; a partir de estos valores, los utilitaristas decían que la utilidad de K era el doble que la de Z, o que generaba el doble de placer. Pareto en cambio consideró que solo era importante conocer la relación de preferencia, es decir, que se prefería K a Z, siendo carente de importancia la cuantía o magnitud en que K se prefería sobre Z. Esta tesis fue aceptada por la

economía estándar, dando origen a la moderna teoría de la utilidad.

No obstante, el concepto de utilidad como cualidad hedónica se ha reestablecido hacia fines de siglo XX, definiéndola como *experienced utility*: “la medida de la experiencia hedónica de un resultado” [Kahneman, 1994]. Los propósitos de su reincorporación al ámbito económico han sido variados: en principio se ha buscado hacer testeable la hipótesis de maximización de utilidad, es decir, convertirla en una proposición empírica [Kahneman et al, 1997]. Por otro lado, Kahneman y Sugden [2005] propusieron reemplazar la idea de *decision utility* (o utilidad entendida como la ponderación o peso asignado a un resultado en una decisión, noción comúnmente utilizada por la moderna teoría económica) por *experienced utility*, a los efectos de encontrar un estándar de evaluación de políticas, debido a que el uso de *decision utility* omitía numerosos patrones de comportamientos, tales como el llamado *focusing illusion* (o exageración de importancia relativa a eventos en los cuales se focaliza la atención), *durability bias* (o sobre-valoración en la durabilidad de cambios en felicidad) o incluso *framing effects* (cambios en las preferencias a partir de cambios en la descripción de

resultados). Finalmente, Kahneman y Krueger [2006] hicieron uso de *experienced utility* para medir el bienestar subjetivo de la comunidad, proponiendo el uso de un índice que exprese la proporción de tiempo que los individuos utilizaban en estados emocionales desagradables: el U-index.

Cabe destacar que los avances en *experienced utility* también han permitido superar el inconveniente de medir la cantidad de placer-dolor de cada uno de los individuos, y para ello se ha apelado a diversos mecanismos psicológicos. Métodos tales como *Day reconstruction Method* desarrollado por Kahneman, Kreuger, Schkade y Schwarz [2004] y *Experienced sampling methodology* [Csikszentmihalyi, 1990; Stone et al, 1999], brindan una alternativa para la medición de *experienced utility*.

Bentham y el utilitarismo.

Bentham introdujo el concepto de utilidad como medida del placer y dolor. Según su tesis, el propósito de las personas radicaba en maximizar el placer, minimizando el dolor, y todos los actos de los individuos se hallaban vinculados a esta idea. De forma tal que el valor de todo ente dependía de

sus cualidades hedónicas. Por ejemplo, Bentham [1780] comentaba que: “el valor de un terreno era valorable en base al placer de todo tipo que le permitía a un hombre hacer uso del mismo y, de manera semejante, los dolores de todo tipo que le evitaba”.

La teoría de Bentham tenía finalidad de jurisprudencia y legislación; más específicamente, el legislador debía buscar el máximo placer para toda la sociedad, a un mínimo dolor posible. El problema a ello se ubicaba en como poder medir esos valores hedónicos. Para ello, Bentham propuso un método de cálculo de la felicidad, estudiando en primer lugar las variables de las cuales dependía la misma. Estas variables eran:

- 1) Intensidad (del estímulo o magnitud de experiencia hedónica por unidad de tiempo)
- 2) Duración (del episodio)
- 3) Certeza o incerteza
- 4) Proximidad o lejanía

Si ahora el propósito era estimar la tendencia de estados hedónicos subsiguientes, debían considerarse dos variables más:

5) Fecundidad (o probabilidad de que a un episodio hedónico le sigan sensaciones del mismo tipo, es decir, placeres si el episodio era placentero, y dolores si era un dolor)

6) Pureza o probabilidad (que no le sigan sensaciones del tipo opuesto)

Finalmente agregaba una nueva variable relacionada con la cantidad de personas implicadas:

7) Extensión (o número de personas involucradas).

Posteriormente, Bentham planteó un método de medir los estados hedónicos a nivel social. El cálculo se basaba en una suerte de "votación", donde las preferencias de los individuos eran consideradas iguales para cada uno de ellos. El interés del legislador no era observar tanto la **cantidad** total de placer y dolor que una política ocasionaba a cada uno de los miembros de la comunidad, sino si dicha política o acto era aceptada o rechazada por los mismos, en base al cálculo del

placer-dolor que generaba la aplicación de ese acto y al resultado neto (o balance) de sumar todos los placeres y dolores ocasionados en una misma persona; luego se sumaban las decisiones de cada uno de los individuos y se evaluaba la conveniencia de la política a aplicar. Bentham [1780] desarrollaba su método de la siguiente manera²:

1.- (Calcular) *El valor de cada placer distinguible que parezca haber sido producido por el acto en primer lugar.*

2.- *El valor de cada dolor que parezca haber sido producido por él en primer lugar.*

3.- *El valor de cada placer que parezca haber sido producido por él después del primero. En esto consiste la fecundidad del primer placer y la impureza del primer dolor.*

4.- *El valor de cada dolor que parezca haber sido producido por él después del primero. Esto constituye la fecundidad del primer dolor y la impureza del primer placer.*

² El fragmento en cursiva corresponde a un extracto de Bentham: *An Introduction to principles of Moraland legislation*, 1780.

5.- Sumar los valores de todos los placeres por un lado y los de los dolores por otro. El balance, si se decantara del lado del placer, daría una buena tendencia global del acto con respecto a los intereses de aquella persona individual; si se decantara del lado del dolor, la mala tendencia global del mismo.

6.- Contabilizar el número de personas cuyos intereses están implicados y repetir el proceso anterior respecto a cada una. Sumar las cifras que representen los grados de tendencia buena del acto con respecto a cada individuo para quien la tendencia global es buena; hacerlo otra vez con respecto a cada individuo para quien la tendencia global es mala. Hacer el balance, que si está del lado del placer, dará la buena tendencia general del acto con respecto al número total o a la comunidad de individuos implicados; si está del lado del dolor, dará la mala tendencia general del mismo con respecto a la misma comunidad.

Podemos ver entonces ciertas particularidades del método utilizado por Bentham:

- En el cálculo debían tenerse en cuenta todas las variables relevantes.

El cálculo del placer-dolor aplicado a un individuo en particular se basaba en estimar si la aplicación de una política determinada ocasionaba bienestar o malestar a dicho individuo. Se consideraba que debían sumarse los placeres de un lado y los dolores del otro, ya que en primer lugar una política económica aplicada podía tener tanto la capacidad de generar bienestar como malestar. A modo de ejemplo, si la política a emplear fuese la aplicación de un impuesto para la limpieza de un río, el individuo sentiría (en el contexto de dicha teoría) placer por tener limpio al río, pero dolor o desagrado por el hecho de pagar el impuesto. Apelando a la subjetividad de las personas, algunas estarían dispuestas a aceptar dicha medida, ya que el placer ocasionado era superior a su malestar. Otras en cambio, expresarían cierta renuencia a la aplicación de la política, en tanto el dolor superaba al placer generado. Más aun, el placer y dolor no radicaban solo en estos dos acontecimientos: si hacemos uso de las variables fecundidad y pureza, podríamos ver que a ciertos episodios le corresponderían otros episodios subsiguientes, los cuales podrían ser tanto placenteros como

dolorosos. Siguiendo con el ejemplo, una persona experimentaría malestar por pagar el impuesto, y placer por el río limpio, siendo el primero mayor al segundo. Pero supongamos que con el río sucio el individuo se encontraba propenso a múltiples enfermedades; esto era también causa de dolor o malestar, y dicha cuantía también debía sumarse a la determinación del valor total. Esta como otras actividades sujetas a valoración debían ser tenidas en cuenta a la hora de hacer el balance de si el acto era en términos netos placentero o doloroso.

- Las preferencias se sumaban imparcialmente.

Luego de hacer el cálculo total, se especificaba si el acto era generador de placer o de dolor. Empero, la **cuantía** total de hedonismo no se tenía en cuenta; solo se denotaba si la política a realizar era aceptada o rechazada por los individuos. Cuando el legislador debía decidir su aplicación, valoraba solo las decisiones de los individuos, mas no la cantidad de placer o dolor. Por ejemplo, consideremos tres individuos tales que a los dos les resulte favorable el pago de impuestos para la limpieza del río y solo al tercero le desagrada. Supongamos también que los grados de

satisfacción (en el supuesto que sean mensurables y comparables entre sí) son distintos: al primer individuo, la utilidad o placer que da la aplicación de esa política es de 8 y la del segundo 2; para el tercero, la utilidad es negativa (porque expresa desagrado) siendo el monto igual a -11. Si definimos a la utilidad total como la suma de utilidades de todos los individuos, entonces está será igual a -1. Si las decisiones se tomasen en relación a la utilidad total, entonces dicha política no se aplicaría. No obstante, esta utilidad total no ha de ser evaluada por el legislador, sino solo las decisiones de aceptación o rechazo en cuanto al acto en cuestión. En cuyo caso, la decisión del legislador hubiese sido la de aplicar la política, ya que era aceptada por la mayoría de la población (2/3 partes).

- Mayor placer para mayor número de personas

Un supuesto fundamental en el sistema de Bentham era el de buscar el máximo placer, haciendo mínimo el dolor. Tal como él comentaba, el legislador en sus decisiones debía de procurar el mayor bien para el mayor número de personas involucradas. De este modo, si la instauración de una política era rechazada por la mayoría de la población, el legislador no

podía aplicarla, debido a que el dolor o malestar era superior al placer generado.

El abandono de la utilidad experimentada.

La moderna teoría económica ha reemplazado el término *experienced utility* (o utilidad en el sentido de Bentham) por el de *decision utility*, el peso de un resultado en una decisión. En este sentido, Kahneman et al [1997] comentan que “la utilidad es inferida de elecciones observadas y alternadamente se utiliza para explicar dichas elecciones”. *Decision utility* es entonces una medida de utilidad ligada al concepto de preferencia, y su uso surge de la imposibilidad de aplicar *experienced utility* al ámbito económico. Existen dos argumentos principales a favor de este reemplazo. Uno es que las experiencias hedónicas no pueden ser medidas u observadas. Tanto la escuela utilitarista como la marginalista sostenían que la utilidad de los bienes era medible así como era medible el peso de los cuerpos. Supongamos dos bienes X y Z, tales que la utilidad de X era de 10 unidades y la de Z 20 unidades. Ambas escuelas asumían que la utilidad de X

era el doble de la de Z; si consideramos utilidad como medida de cualidad hedónica, entonces X daba el doble de placer que Z. Sin embargo, no se encontró ninguna forma de medir las cantidades estas experiencias hedónicas.

El segundo argumento se halla relacionado con el postulado de racionalidad, unido a una redefinición del término utilidad. En esta nueva concepción, la utilidad se desvincula de todo significado sensorial: si la utilidad de X es mayor a la de Z, eso no significa que el placer que da X es superior del de Z; solo muestra que si a un individuo se le presentase esa opción, preferirá X a Z. El supuesto de racionalidad implica que el individuo conoce todas las alternativas posibles que se le presentan, que ellas se encuentran sujetas a valoración por parte de dicho individuo, y que las ordena y clasifica en relación a un orden de preferencias. Cabe destacar que estas valoraciones representan el orden de dichas preferencias. Si la utilidad de X es de 20 unidades y de 10 la utilidad de Z, esto solo significa que X se prefiere sobre Z. Bajo un argumento semejante, Kahneman y Sugden [2005] consideran que el abandono de la utilidad como cualidad hedónica por el de utilidad como reflejo de preferencias tiene su causa en la economía normativa: “en la teoría neoclásica

del bienestar, la medida de bienestar para cualquier individuo dado es su utilidad. Si utilidad es interpretado como decision utility y si al agente económico racional se lo supone maximizador de utilidad, entonces no hay necesidad de preocuparse acerca de si están eligiendo aquellas cosas que disfrutarán al consumir” [Kahneman y Sugden, 2005]. No obstante, investigaciones del lado del *Behavioural Economics* cuestionaron el supuesto de racionalidad, y mostraron que los agentes económicos presentan de hecho racionalidad acotada. Esto genera inconvenientes en la medición de la decision utility, ya que el individuo no revela coherentemente sus preferencias. Kahneman y Sugden proponen que ante dicha imposibilidad de medición, debería de utilizarse el concepto de experienced utility. Al respecto, presentaron una serie de argumentos que muestran la necesidad de dicho reemplazo:

(1) Actitudes versus preferencias.

Las respuestas de los individuos a determinadas preguntas se pueden categorizar bajo dos conceptos: actitudes y preferencias. La idea de preferencia es comúnmente utilizada en el ámbito económico, donde lo que se hace es

comparar los distintos estados alternativos; se considera por tanto que las preferencias son *extensionales*. Las actitudes en cambio se utilizan en la psicología, y su característica es la no-extensionalidad: lo que se reportan son eventos en la mente.

Asimismo, la teoría estándar de la decisión asume que las preferencias de los individuos permanecen inalteradas ante cambios en la descripción de resultados, supuesto que se ha denominado *extensionalidad* [Arrow, 1982] o *invariancia* [Tversky y Kahneman, 1986]. Empero, en Tversky y Kahneman [1981] se mostró que los agentes aceptan de manera pasiva la formulación dada, lo cual viola el supuesto de extensionalidad y por ende la lógica de las preferencias. “La invariancia no se cumple en los *Framing Effects* o *efectos marco*, en los que descripciones equivalentes extensionalmente hablando llevan a elecciones diferentes, al modificar la notabilidad relativa de los diferentes aspectos del problema” [Kahneman, 2003]. En cambio, como las actitudes no son extensionales, son susceptibles de framing effects, lo cual permite que no se viole la lógica de las actitudes [Kahneman y Sugden, 2005]. Además, el dominio de las actitudes es más amplio que el de las preferencias: un

individuo puede tener actitudes de varias cosas y no por ello tener que preferir una a otra, puesto que no existen alternativas para elegir.

(2) Incapacidad de predicciones correctas:

Mientras la decision utility corresponde a un concepto ex ante (refleja las preferencias del individuo antes del proceso de elección), experienced utility es una noción ex post, esto es, representa la experiencia hedónica que resulta de episodios acontecidos. Para que los valores de decision utility se asemejen a los de experienced utility, se requerirá que los individuos sean capaces de hacer razonablemente predicciones precisas de las consecuencias de sus actos; en otras palabras, deben ser buenos predictores. Ahora bien, existen investigaciones que muestran que los individuos fallan en sus predicciones. Kahneman y Sugden consideran que la mala predicción en lo afectivo tiene su causa en dos razones: **(a)** diferencia entre estados y transiciones, **(b)** mecanismos de adaptación.

(a) es una noción derivada de Prospect Theory [Kahneman y Tversky, 1979], en la cual se plantea que los sujetos no

valoran los estados finales de una dotación, sino los cambios o variaciones patrimoniales a partir de una posición activa dada. De modo tal que la percepción depende de la referencia: “los atributos percibidos de un estímulo focal reflejan el contraste entre dicho estímulo y un contexto de estímulos previos y concurrentes” [Kahneman, 2003]. Por ejemplo, la brillantez percibida depende de la luminosidad del entorno por el cual se percibe el estímulo. Y puesto que los individuos perciben las transiciones y no los estados, luego no tendrán una noción clara de lo que se siente SER parapléjico, aunque sí sobre lo que sería VOLVERSE parapléjico.

En **(b)**, los acontecimientos de felicidad o infelicidad se vuelven cada vez menos intensos, puesto que el individuo se va adaptando o acostumbrando a aquellos estímulos generadores de felicidad o infelicidad. Dentro de los mecanismos de adaptación podemos citar el de Reorganización de la Atención (redeployment of attention): una vez que un acontecimiento o estímulo es percibido por el individuo, este se acostumbra a dicho estímulo, por lo que refocaliza la atención a nuevos estímulos. Por ejemplo, los niveles de bienestar reportados por un parapléjico son

bastante mayor que los que una persona sana podría esperar (o estimar). La explicación a ello es que los individuos redireccionan la atención. Esto es, no piensan permanentemente en su condición de parapléjico, sino que lo asimilan, por lo que terminan redireccionando su atención a nuevos estímulos. A partir de este mecanismo podemos inferir que los individuos realizan predicciones incorrectas respecto de sus estados hedónicos futuros. Esta falla ha sido conocida como “Durability bias”: los individuos suelen sobreestimar la duración de los cambios de felicidad que siguen a la transición de nuevas circunstancias objetivas [Gilbert et al, 1998]. Un experimento desarrollado por Gilbert et al [1998] en Texas en torno a las elecciones para gobernador durante 1990, mostró que los individuos presentarían descontento si su candidato perdía las elecciones. Empero, un mes después de dichas elecciones, cuando se volvieron a hacer preguntas acerca del bienestar, los resultados de la elección no tuvieron aquellos efectos que se pensaban. También es destacable el mecanismo de *Focusing illusion* (o ilusión focal): los parapléjicos se “acostumbran” a ser así, por lo que redireccionan la atención; no piensan que lo son. En cambio, los individuos sanos cuando piensan en los parapléjicos, consideran que estos están pensando que son parapléjicos,

lo cual es un error. Por tanto, cuando a un individuo se le pide que piense en algo, este focaliza su atención en ello, por lo que tiende a exagerar su importancia relativa [Kahneman y Sugden, 2005].

Particularidades de experienced utility

Definiendo a *experienced utility* como una cualidad hedónica, corresponde ahora categorizar los reportes de esta: específicamente, *experienced utility* puede ser reportada en tiempo real o en evaluaciones retrospectivas de episodios pasados [Kahneman et al, 1997]. A la primera clase de reporte se la ha denominado *instant utility*; a la segunda, *remembered utility*. *Instant utility* corresponde a la medida de placer o dolor en el mismo momento en que se percibe el estímulo. *Remembered utility* en cambio es la evaluación retrospectiva de un resultado pasado extendido temporalmente, o en otros términos, la valoración presente de una experiencia hedónica pasada. Por otra parte, las experiencias hedónicas son señales biológicas percibidas por los individuos: para el caso del placer, existe una señal de “siga” o de guiar al organismo a continuar la actividad. El

dolor en cambio es una señal de “pare”, lo cual guía al organismo a interrumpir la actividad.

En Kahneman et al [1997] se propuso restablecer el concepto de utilidad como sinónimo de cualidad hedónica. Una premisa que asumieron fue que los estados hedónicos podían ser medidos a partir de ciertos instrumentos utilizados por la psicología: la medida tanto de experiencias subjetivas como la estimación de funciones que explican distintos comportamientos individuales en relación a estímulos presentes y pasados son tópicos bien establecidos en el campo de investigaciones psicológicas [Stevens, 1975; Wegener, 1982; Kahneman et al, 1997]. Puede considerarse que *experienced utility* tiene mucho en común con la temperatura subjetiva: en esta existen expresiones de calor o frío, por lo que logramos tener escalas separadas para dichos atributos: una escala que exprese la sensación de calor, y otra la de frío. Estas escalas son compartidas por un punto neutral, que provee un punto cero natural para dimensiones bipolares [Kahneman et al, 1997]. Dicho punto representa una sensación de “ni calor ni frío”. Para el caso de las experiencias hedónicas, el caso es análogo: existe un punto neutral, que expresa “ni placer ni dolor” para una sensación, y

a partir de este punto se obtienen dos escalas, una para el placer, otra para el dolor.

A su vez, los reportes de valores hedónicos pueden obtenerse a través de varias expresiones y de distintas escalas. Una de las formas de medir instant utility es proveer a los individuos una escala numérica (por ejemplo 0-10), bajo la cual los individuos rankean la sensación hedónica correspondiente: cero representaría el punto neutral (ni placer ni dolor) y 10 la expresión hedónica más alta posible (máxima expresión de placer o dolor percibida por el individuo). Un ejemplo se ha dado en Redelmeier y Kahneman [1996], donde a una serie de participantes se los sometió a una colonoscopia: los individuos debían reportar su nivel de discomfort en una escala 0-10. Otra variante de medir experienced utility se ha visto en Gottman y Levenson [1985]: en el experimento se pasaron un conjunto de videos, tanto placenteros como desagradables, y los individuos debían de valorar la experiencia hedónica correspondiente. La forma de medirla fue a través de una palanca móvil que contenía un conjunto de luces de colores para que ellos sean provistos de una medida continua de experienced utility durante la presentación de cada filmación. Además, los valores

hedónicos podían ser suplementados por indicadores psicológicos de intensidad y calidad emocional, por ejemplo, incluyendo medidas objetivas de expresiones faciales [Frank, 1988].

Jevons [1871] comentaba que todo sentimiento, sea de placer o de dolor, poseía dos variables fundamentales: la duración y la intensidad. Todo sentimiento había de durar algún tiempo, y mientras duraba, podía ser más o menos intenso. Para una misma cantidad de tiempo, si se incrementaba la intensidad de un placer, la utilidad resultante también se vería incrementada. Por otro lado, “si la intensidad de un sentimiento permaneciera constante, la cantidad de sentimiento aumentaría con su duración” [Jevons, 1871]. Más aun, Jevons afirmaba que la relación entre duración y utilidad era directamente proporcional: si se duplicaba la duración, para un mismo grado de intensidad, el placer o la utilidad sería también el doble. Podemos ver entonces que en la tesis de Jevons el tiempo es una variable determinante de experienced utility o utilidad en el sentido de Bentham. No obstante, investigaciones de laboratorio han constatado que la duración de un episodio no tiene relevancia en la determinación de remembered utility. Por ejemplo, en Mowrer

y Solomon [1954] se observó que el miedo en ratas que fueron expuestas repetidamente a shocks eléctricos guardaba una alta correlación con la intensidad del shock, pero no con la duración. En Kahneman [1994] al estudiar el rol central de la duración como determinante de la utilidad de los episodios, se obtuvieron dos generalizaciones empíricas: **(1)** Peak-End evaluation y **(2)** decisiones por remembered utility.

En **(1)** el valor de remembered utility de una experiencia hedónica (sea placentera o dolorosa) se predice con precisión promediando el “peak” o valor de mayor intensidad de la instant utility y la instant utility reportada cerca del final de la experiencia [Kahneman, 1994]. En **(2)**, cuando a los individuos se les da la opción a elegir entre la repetición de diversos episodios, ellos generalmente eligen aquel que posee remembered utility más alta [Kahneman et al, 1997]. Dos consecuencias de peak-end han sido confirmadas: **(i)** olvido de la duración (duration neglect) y **(ii)** violación de la monotonidad. **(i)** muestra que la duración de una experiencia tiene un efecto poco significativo o incluso nulo en la determinación de remembered utility: la evaluación retrospectiva de todo el placer o dolor total no es afectado por la duración de un episodio [Kahneman, 1994]. Esto es un

caso particular del concepto de Extension Neglect: “manteniendo las demás cosas igual, un incremento en la extensión de una categoría aumentará el valor de los atributos extensionales, pero no modificará los valores de sus atributos prototípicos” [Kahneman, 2003]. (ii) hace referencia a que el valor de remembered utility de un episodio averso puede ser reducido si se añade un período extra de discomfort que reduzca el promedio del peak-end.

En Redelmeier y Kahneman [1996a] se confirmó la tesis de Duration Nelect y Peak End evaluation, a partir de una colonoscopia realizada a pacientes de un hospital de Toronto. En dicho estudio, la duración del procedimiento varió entre los 4 y los 69 minutos en una muestra de 101 pacientes. Sorpresivamente, estas variaciones en la duración no tuvieron efectos significativos en evaluaciones retrospectivas [Kahneman, 1994]. En lugar de eso, las valoraciones dependían de la intensidad del “peor” dolor (el dolor más intenso) y por la intensidad de discomfort en los últimos minutos del procedimiento. La correlación entre el promedio Peak-End con el grado global del procedimiento fue de 0,67, mientras que la correlación entre la evaluación global y la duración del procedimiento fue solo de 0,03. La violación de

la monotonicidad también fue confirmada: en Redelmeier y Kahneman [1996b] testearon dicha hipótesis con una muestra de 682 pacientes sometidos a una colonoscopia. El resultado fue que prolongando la duración de dicha colonoscopia se mejoraron las evaluaciones retrospectivas globales. Cabe aclarar que esta extensión temporal implicaba una intensidad más baja, lo que redujo el promedio del Peak-End. La violación de la monotonicidad de hecho se observó en Redelmeier et al [2003], también a través de un estudio de colonoscopia. A la mitad de los pacientes no se les quitó el aparato inmediatamente, sino que el médico esperó cerca de un minuto, sin mover el aparato. Durante el período extra se experimentó una sensación desagradable, pero se garantizó que el procedimiento nunca terminara con un dolor severo [Kahneman, 2003]. A la hora de hacer una evaluación global, los resultados fueron en estas condiciones más favorables que en las condiciones de control [Redelmeier et al, 2003].

La violación de la monotonicidad (también llamada violación de la preponderancia) se vio incluso reflejada en las elecciones. En Kahneman et al [1993] se realizó un experimento en el cual a los participantes se les pidió que introduzcan la mano en agua fría. Se realizaron dos pruebas:

un episodio corto (60 segundos) en el cual la temperatura del agua era de 14°C, y un episodio largo (90 segundos) en el que el agua se fue calentando progresivamente; los primeros 60 segundos a 14°C, los últimos 30 segundos a 15°C. La valoración de remembered utility fue superior en el episodio más largo. La violación de monotonicidad temporal y duration neglect también se observó en Scheiber y Kahneman [1996], en el cual se estableció como estímulo aversivo la aplicación de sonidos fuertes. Se realizaron dos pruebas: en una se emitían sonidos desagradables a 78 decibeles, en la otra se emitieron los mismos sonidos pero a una intensidad de 66 decibeles, pero de mayor duración. A la hora de repetir el ensayo, los individuos tuvieron preferencia significativa por el de mayor duración.

Kahneman sostiene que cuando se realiza una elección, se elige la alternativa que se asocia con la utilidad recordada más alta [Kahneman, 2003]. Para cualquiera de los experimentos recientemente mencionados, los individuos que fueron sometidos a estos tienen un recuerdo de los episodios. Pero en este recuerdo, los sujetos no gastan mucho tiempo en comparaciones analíticas antes de elegir: ellos simplemente piensan cual de los recuerdos les disgustaba

menos, por lo que las elecciones se hicieron comparando los distintos valores de remembered utility. Estos episodios se representan en la memoria a través de un momento característico, y que el episodio sea placentero o desagradable es algo que viene determinado por la utilidad recordada de ese momento [Kahneman, 1994; Kahneman, 2003]. La maximización de la utilidad se convierte ahora en maximizar remembered utility: a la hora de repetir un episodio, los individuos van a elegir aquel episodio (placentero) de remembered utility más alta, o bien aquel episodio (doloroso) de remembered utility más baja.

La medición Experienced Utility para Políticas Económicas.

Kahneman y Sugden [2005] exploraron la posibilidad de basar la evaluación de políticas económicas bajo la medición de experienced utility en lugar de decision utility. Ellos se preguntaron de qué modo podía medirse la experienced utility, y comentaban: “necesitamos un método de medición el cual obtenga información acerca de estados actuales de experiencias hedónicas, no actitudes sobre determinadas cuestiones o respuestas afectivas sobre transiciones”

[Kahneman y Sugden, 2005]. Focusing Illusion generaba inconvenientes a la hora de medir decision utility; por tanto, se necesitaba de un método por el cual se pudiese evitar focusing illusion. Además, se necesitaba generar información en una forma que permitiera identificar o predecir los efectos hedónicos de determinadas políticas.

Se han diseñado varios métodos para medir experienced utility, entre los cuales podemos destacar: **(a)** *Moment-based happiness* [Kahneman, 2000a]; **(b)** *Experienced sampling methodology* [Csikszentmihalyi, 1990; Stone et al, 1999]; **(c)** *Day Reconstruction Method* [Kahneman, 2004].

En **(a)** se procura medir la calidad de la experiencia hedónica de un episodio. Supongamos que deseamos medir la felicidad de personas ciegas: no se les preguntará acerca de la felicidad total en sus vidas, así como tampoco se les pedirá que imaginen volverse ciegos para evaluar que efectos tendrían sobre su felicidad. Por el contrario, este método propone medir la calidad de la experiencia hedónica que las personas ciegas están teniendo momento a momento, en el curso de sus vidas. Similarmente, si se desea descubrir si una calle arbolada hace más feliz a la gente, entonces se

medirán las experiencias de las personas mientras camine tanto por una calle arbolada como por una calle sin árboles [Kahneman y Sugden, 2005; Kahneman, 2000a].

Experienced sampling methodology (b) fue desarrollado para recolectar información sobre los sentimientos reportados por los individuos en determinados momentos del día. Básicamente, se trata de un método por el cual cada participante lleva consigo una palmtop (computadora de bolsillo) que emite un sonido en momentos aleatorios a lo largo del día. En ese preciso momento, cada sujeto debe responder a cuestiones acerca de su situación actual y estados afectivos [Kahneman y Sugden, 2005]. A los participantes se les presenta un menú en el cual deben especificar su localización física, la actividad realizada justo antes de la emisión del sonido, y los individuos con los que se encontraba interactuando. También se les pregunta en relación a su experiencia subjetiva actual, indicando el grado en el cual sienten la presencia o ausencia de varias sensaciones, tales como alegría, tristeza, cansancio, impaciencia, etc. [Kahneman y Krueger, 2006].

(c) Puesto que este método es difícil de implementar en muestras grandes de población, Kahneman et al [2004] propusieron el *Day Reconstruction Method* (DRM), el cual ha sido utilizado en estudios de felicidad. A los individuos se les pregunta acerca del día anterior, y se les pide que lo descompongan en episodios cortos (por ejemplo, en la hora del almuerzo, de viaje volviendo del trabajo, etc.). “Luego ellos describen cada episodio indicando: cuando un episodio empieza y cuando termina, que estaban haciendo (seleccionando una actividad de una lista que se les proveía), donde estaban, y con quien se encontraban interactuando” [Kahneman y Krueger, 2006]. Para establecer como ellos se sentían emocionalmente en cada episodio, se les pedía que reporten la intensidad de sus emociones a lo largo de una serie de categorías sobre una escala de 0-6, donde cero representaba una magnitud nula de satisfacción (ni placer ni dolor) y 6 la intensidad máxima de afección emocional. A su vez, las categorías eran: impaciente, feliz, frustrado/molesto, deprimido, competente/capaz, cálido/amigable, enojado, hostil/enfadado, aburrido, gozando, crítico y cansado [Kahneman et al, 2004]. En Kahneman y Krueger [2006] se observó que los resultados en un estudio de “cansancio a lo largo del día” en una muestra de 909 trabajadoras de Texas

fueron muy similares aplicando tanto el Day Reconstruction Method como el Experienced Sampling Method.

Un problema que subyace a esta clase de medición es la relatividad de las valoraciones: la ponderación de un episodio hedónico se hace bajo una escala “determinada” por el sujeto mismo, de forma tal que las magnitudes de experienced utility son inconmensurables para los distintos individuos. Verbigracia, supongamos que se desea medir el bienestar dos sujetos A y B. Mientras A dice ser muy feliz, B argumenta que solo es feliz. Sería falaz decir que A es más feliz que B, ya que la forma de valorar de uno **no** va a ser necesariamente semejante a la del otro. Posiblemente, A sea una persona que magnifique los episodios, que utilice frecuentemente superlativos. Así, a la hora de responder por la intensidad de una emoción particular, puede que A responda 6 mientras B responda 4, y no obstante el 4 de B sea equivalente al 6 de A. Bajo esta consideración, es cuestionable dar una interpretación cardinal a los valores numéricos asignados por los individuos en relación a sus estados emocionales o satisfacción en la vida a causa del potencial uso de escalas personales [Kahneman y Krueger, 2006]. En este mismo trabajo se propone un índice, al cual

denominaron U-index, el cual mide el nivel de desagrado y permite superar dicho problema. El U-index es una medida ordinal de los niveles de emoción, donde se calcula la proporción de tiempo que un individuo pasa en eventos desagradables. El propósito de dicho índice subyace en esquematizar una posible medida del bienestar subjetivo.

El primer paso a la hora de computar el U-index es determinar si un episodio es placentero o desagradable. Kahneman y Krueger comentan: “un episodio como desagradable si el momento de mayor intensidad reportado para un episodio, es decir, si la clasificación más alta sobre cualquiera de las dimensiones negativas de afecto es estrictamente superior que el máximo de clasificación de las dimensiones positivas del afecto” [Kahneman y Krueger, 2006]. Como se trata de un ranqueo puramente ordinal de los sentimientos de cada episodio, se vuelve irrelevante que A utilice la porción 0-4 de la escala 0-6 de intensidad y que B utilice todo el rango.

Una vez que se hayan clasificado los episodios como placenteros o desagradables, se mide el U-index como fracción del tiempo que se gasta en estados desagradables.

El U-index se computa para cada individuo (proporción del tiempo que esta persona se encuentra en un estado emocional desagradable), y luego se promedia para una muestra de individuos. El mismo índice puede ser utilizado para describir situaciones o actividades. Por ejemplo, en Kahneman y Krueger [2006] se calculó el U-index para las distintas actividades que realizaban una muestra de 909 trabajadoras de Texas. A partir de esto se encontró que en promedio, el 28 por ciento del tiempo que utilizaban viajando todas las mañanas al trabajo era desagradable, mientras que el 21 por ciento del tiempo en el que trabajaban era desagradable, y solo un 4 por ciento en relaciones íntimas. Los U-index para cada actividad eran respectivamente 0,28 para el viaje mañanero al trabajo, 0,21 para trabajar y 0,04 para relaciones íntimas. También se calculó que en promedio estas mujeres gastaban aproximadamente un 17 por ciento de su tiempo total en estados desagradables, que el U-index caía ligeramente con el ingreso doméstico (18,9 por ciento para personas con ingresos menores a U\$S 35000 y 16,6 por ciento para ingresos mayores o iguales a U\$S 55000). El U-index también guardaba una relación inversa con la edad: aquellas personas de entre 18 y 24 años, utilizaban aproximadamente un 22 por ciento en estados

desagradables, comparado con 19 por ciento para aquellos de entre 25 y 44 años y un 15 por ciento para edades de entre 45 y 64. La personalidad se consideró una variable importante en la determinación de los valores afectivos: las personas que sufrían de estados depresivos utilizaban mayor parte del tiempo en situaciones desagradables que aquellas que no lo eran.

A los efectos de reducir el monto de tiempo utilizado en estados emocionales desagradables, Kahneman y Krueger propusieron distintas soluciones. La primera focalizaba en intervenciones de salud mental (tales como suministrar drogas anti-depresivas) para aquellas personas que gastaban gran parte de su tiempo en estados desagradables. La segunda focalizaba en la asignación del tiempo. Por ejemplo, aquellas intervenciones que redujesen el monto del tiempo utilizado en viajar al trabajo (tales como subsidios al uso compartido de automóviles) podrían tener efectos beneficiosos en los estados emocionales de los individuos.

El sentido pragmático del U-index subyace en una propuesta de la medición del bienestar social, basado en una unidad de medida objetiva: el tiempo. La naturaleza del U-index permite

comparaciones del nivel de bienestar entre países, aunque puede verse distorsionado por diferencias culturales o lingüísticas a la hora de responder a preguntas específicas sobre estados emocionales.

Limitaciones

El intento de restablecer el concepto de utilidad como cualidad hedónica ha venido desde el lado de la psicología, y ha permitido abrir el campo a nuevas investigaciones empíricas, tales como convertir la hipótesis de maximización en una proposición empírica, evaluar los cambios hedónicos en los individuos ante la aplicación de políticas económicas, medir el bienestar subjetivo, etc.

Se ha visto que la noción de *experienced utility* encierra dos conceptos de íntima vinculación: *instant utility* y *remembered utility*. Los valores de *instant utility* se obtienen en relación a dos variables: la extensión en el tiempo (o duración de un episodio) y la intensidad de la experiencia hedónica. Esto muestra que todo episodio, sea placentero o doloroso, se extiende en el tiempo, y a lo largo de este, se experimentarán distintas sensaciones de intensidad emocional. Por su parte,

Remembered utility ha sido definida como la evaluación retrospectiva de un episodio pasado. De esto se sigue necesariamente que experienced utility es un concepto ligado a la existencia de un episodio en donde se suceden distintas experiencias hedónicas, o en otros términos, los valores de experienced utility se pueden obtener solo de episodios extendidos temporalmente.

Se puede objetar al **uso general** de experienced utility las limitaciones que encierran su propia definición, ya que no todo ente que posea utilidad se va a presentar en un episodio hedónico: una ganancia de \$100 será valorada positivamente por el individuo, es decir, tendrá esto una utilidad positiva. Empero, esta ganancia no puede representarse en un espacio temporal, y menos aun en intensidades variables a lo largo del episodio. Tampoco es aplicable al ámbito de las valoraciones en tanto satisfacción de necesidades. Por ejemplo, en Menger toda entidad poseía utilidad en la medida que era útil al hombre, y era útil cuando tenía la capacidad de satisfacer una necesidad de este. Cuando se daban dichas circunstancias, la entidad en cuestión era considerada como un bien. Es evidente que cuando dicha entidad no satisficiera más ninguna necesidad, se convertiría en algo inútil, y por

tanto de utilidad nula. Así, los valores de utilidad se hallaban vinculados a la medida en que un bien era útil para el individuo, no si este generaba placer. No obstante, podríamos decir que la economía conductual no ha permanecido ausente a esta limitación. Por ejemplo, en Kahneman y Sugden [2005] se cita un caso bajo el cual las personas podían estar dispuestas a pagar impuestos que financien la conservación de especies que probablemente nunca vean. Esto podía entenderse mejor como la expresión de una actitud moral o política más que como un significado de comprar una experiencia hedónica.

BIBLIOGRAFIA

Arrow, K. J. (1982): *Risk Perception in Psychology and Economics*, Economic Inquiry, vol. 20, n° 1, pp. 1-9.

Bentham, J. (1789): *An Introduction to the Principle of Morals and Legislations* (1789); reprinted (Oxford, UK: Blackwell, 1948).

Csikszentmihalyi, M. (1990): *Flow: The Psychology of Optimal Experience*. New York: HarperCollins.

Frank, R. H. (1988): *Passions within Reason: The Strategic Value of the Emotions*, New York: Norton.

Gilbert, D., E. Pinel, T. Wilson, S. Blumberg, T. Wheatley (1998): *Immune Neglect: A Source of Durability Bias in Affective Forecasting*, Journal of Personality and Social Psychology 75, pp. 617–638.

Gottman, J., R. Levenson, (1985): *A Valid Procedure for Obtaining Self-Report of Affect in Marital Interaction*, Journal of Consulting and Clinical Psychology, LIII, pp. 51-160.

Hausman, D. (1992): *The Inexact and Separate Science of Economics*, The philosophical Review, Vol. 103, Nº 2. Cambridge : Cambridge University Press.

Henderson, J., R. Quandt, (1962): *Microeconomic Theory: a mathematical approach*, traducción al castellano en Ed. Ariel, Barcelona.

Jevons, S. (1871): *The Theory of Political Economy*.

Kahneman, D. (1994): *New Challenges to the Rationality Assumption*, Journal of Institutional and Theoretical Economics, CL, pp. 18-36.

Kahneman, D. (2000a): *Experienced Utility and Objective Happiness: A Moment-based Approach*, in A. Tversky and D. Kahneman, eds., *Choices, Values, and Frames*, Cambridge: Cambridge University Press.

Kahneman, D. (2000b): *Evaluation by Moments: Past and Future*, in A. Tversky. D. Kahneman, eds., *Choices, Values, and Frames*, Cambridge: Cambridge University Press.

Kahneman, D. (2003): *Maps of Bounded rationality: psychology for behavioral Economics*, The American Economic Review, Vol. 93, No. 5., pp. 1449-1475.

Kahneman, D., B. Fredrickson, C. Schreiber, D. Redelmeier, (1993): *When More Pain Is Preferred to Less: Adding a Better End*, Psychological Science, IV, pp. 401-05.

Kahneman, Daniel, A. Krueger, D. Schkade, N. Schwarz, A. Stone, (2004): *A Survey Method for Characterizing Daily Life Experience: The Day Reconstruction Method*, Science. 306:5702, pp. 1776–780.

Kahneman. D., A. Krueger, (2006): *Developments in the Measurement of Subjective Well-Being*, Journal of Economic Perspectives—Volume 20, Number 1, pp. 3–24

Kahneman, D., J. Snell, (1992): *Predicting a Changing Taste: Do People Know What They Will Like?*, Journal of Behavioral Decision Making-e., V. pp. 187-200.

Kahneman, D., R. Sugden, (2005): *Experienced Utility as a Standard of Policy Evaluation*, Environmental & Resource

Economics 32. DOI 10.1007/s10640-005-6032-4, pp.161–181.

Kahneman, D., A. Tversky (1979): *Prospect Theory: An Analysis of Decision Under Risk*, *Econometrica* 47, pp. 263–291.

Kahneman, D., C. Varey (1991): *Notes on the Psychology of Utility*, in J. Elster and J. E. Roemer, eds., *Interpersonal Comparisons of Well-Being: Studies in Rationality and Social Change*. New York: Cambridge, pp. 127-63.

Kahneman, D., P. Wakker, R. Sarin, (1997): *Back to Bentham? Explorations of Experienced Utility*, *Quarterly Journal of Economics*, vol. 112, n° 2, mayo, pp. 375-405.

Mowrer, O. L. Solomon, (1954): *Contiguity vs. Drive-Reduction in Conditioned Fear: The Proximity and Abruptness of Drive-Reduction*, *American Journal of Psychology*, LXVII, pp. 15-25.

Redelmeier, D., D. Kahneman, (1996a): *Patients' Memories of Painful Medical Treatments: Real-Time and Retrospective Evaluations of Two Minimally Invasive Procedures*," *Pain*, CXVI, 3-8.

Redelmeier, D., D. Kahneman, (1996b): *Improving the Memory of a Colonoscopy*.

Schreiber, C., D. Kahneman, (1996): *Beyond the Peak and End Hypothesis: Exploring the Relation between Real-Time Displeasure and Retrospective Evaluation*, Princeton University.

Stevens, S., (1975): *Psychophysics: Introduction to its Perceptual, Neural and Social Prospects*, New York: Wiley.

Stone, A., S. Shiffman, M. DeVries, (1999): *Rethinking Self-report Assessment Methodologies*, in D. Kahneman, E. Diener and N. Schwarz, eds., *Well-Being: The Foundations of Hedonic Psychology* (pp. 26–39). New York: Cambridge University Press.

Tversky, A., D. Kahneman, (1981): *The Framing of Decisions and the Psychology of Choice*, *Science*, vol. 211, n° 4481, pp. 453-458.

Tversky, A., D. Kahneman, (1986): *Rational Choice and the Framing of Decisions*, *Journal of Business*, vol. 59, n° 4, pp. 251-278.

Varey, C., D. Kahneman, (1992): *Experiences Extended across Time: Evaluation of Moments and Episodes*," Journal of Behavioral Decision Making, V, pp. 169-86.

Wegener, B., (1982): *Social Attitudes and Psychophysical Measurements*, Hillsdale, NJ: Lawrence-Erlbaum.

Segunda Sección:

Crítica de las nociones
convencionales de racionalidad y
utilidad

DE LA NEUROLOGÍA A LA DECISIÓN RACIONAL: EL CAMINO DEL FUNCIONALISMO HOMUNCULAR.

Alejandro G. Miroli (CBC/FCE UBA. USAL. Instituto de Profesorado A-1309 “Alfredo Palacios”)

Resumen.

En su libro *Consciousness* (Cap. 4º), William Lycan introduce la tesis de la continuidad de los niveles de la Naturaleza (tesis que bautiza **homuncular functionalism**) como una tesis explicativa de las relaciones entre los niveles neural/cerebral y mental/psíquico. Esta teoría puede aplicarse –casi sin cambios importantes- a un dominio teórico que se expande notablemente: la neuroeconomía. Y ello porque los problemas teóricos –ontológicos y epistemológicos- de la filosofía de la psicología, adquieren un valor nuevo en este dominio. Mi interés en esta comunicación es defender la pertinencia de un enfoque del tipo del funcionalismo homuncular para el tratamiento de un problema propio de las teorías de la decisión: la integración de niveles de análisis neurales o sub-agenciales, agenciales y supra-agenciales propios de la neuroanatomía.

Neuroeconomía, decisión, reduccionismo, funcionalismo, niveles de agencia.

I. El problema.

Si el desarrollo de las neurociencias tuvo un impacto en las ciencias económicas, fue en el Marketing donde dicho impacto se desarrolló con mayor fuerza: ya desde los años 90 se empleaban técnicas de Imagen de Resonancia Magnética funcional (fMRI) para el examen de áreas cerebrales de sujetos en situaciones experimentales vg al observar avisos con colores y formas diversas, al escuchar jingles publicitarios, al examinar envases de productos, etc. Tal fue ese desarrollo que en 2002, Ale Smidts –investigador del Erasmus Research Institute of Management (Erasmus University)- acuñó el término *Neuromarketing* para denominar el creciente número de estudios que aplicaban técnicas de mapeo neural para el examen del comportamiento y de la toma de decisiones de los consumidores.

Una forma de entender el tema es tomar algún caso modelo de investigación en neuromarketing v.g. el estudio ya clásico que desarrollo un equipo de investigadores del Baylor College of Medicine (cf. McClure, 2004) –inspirados en una prueba

previa de consumo a ciegas de Coca Cola y Pepsi Cola-en el que se investigo el impacto que tenía la imagen de marca en pruebas de consumo de bebidas cola mientras se registraba la actividad neural asociada de los sujetos experimentales. En ese estudio que comparaban dos grupos de consumidores uno ciego respecto de la marca que bebían y el otro informado de ellas; el equipo hallo que “For the anonymous task, we report a consistent neural response in the ventromedial prefrontal cortex that correlates with subjects-behavioral preferences for these beverages, In the brand-cued experiment, brand knowledge for one of the drinks shad a dramatic influence on expressed behavioral preferences and on the measures brain responses.” (McClure et al, 2004, 379). En este sentido los autores pretendieron avanzar más allá de las investigaciones efectuadas hasta ese momento, investigaciones en las que el énfasis se había puesto en el llamado problema sensorial i.e. “... the perceptual discrimination of odors and flavors, the correlated neural responses, and the modulation of both by non-odor or non-favor stimuli. <ya que reconocían que> there is another large piece of the problem to understand... behavioral preferences are modulated by an enormous number of sensory variables,

hedonic states, expectations, semantic printing, and social context” (idem).

Este párrafo presenta dos nociones claves: **correlación** y **modulación**; la primera es una noción en el análisis de las relaciones lineales entre variables cuya tratamiento formal está estandarizado; en particular eso supone que debe haber unidades de medición y un supuesto de continuidad; pero la correlación no es ni autoriza inferir una relación determinista: la dependencia extensional –correlación- no supone dependencia intensional –causalidad. Por ello la dependencia intensional supone un elemento conceptual agregado: la modulación, una noción compleja, por la que debe entenderse la totalidad de los factores determinantes cualitativamente diferentes –desde las variables neurológicas y los sesgos sensoriales hasta el contexto social- que conforman el complejo etiológico de la toma de decisiones del consumidor y sus principios de conexión. En ese sentido los autores que su investigación supone una novedad, ya que hasta su trabajo “The neural substrates underlying food and drink preferences and their influence by cultural images have to be not been explored” (idem). Y su estrategia general será determinar como “...cultural information can modulate

reward-related brain responses....” y, en ese caso se intentarán examinar aquellas “...visual images and marketing messages that have insinuated themselves into de nervous systems...” (McClure et al, 2004, 380).

Pero ¿Qué significa que la imagen de marca Coca-Cola modula el comportamiento de consumo de bebidas cola? En términos del fin que se proponen McClure & otros será mostrar que “... our results suggest that there might be parallel mechanisms in the brain cooperating to bias preference.” (McClure et al, 2004, ídem). Pero en el marco del estudio encarado, *cooperación* debe entenderse como *correlación*; de ese modo el programa del neuromarketing supone que:

Conjunto de elementos que modulan el comportamiento molar = Conjunto de cooperaciones/correlaciones psicofísicas.

Pero esta identidad es falsa; primero porque (para Φ = la corteza dorsolateral prefrontal/estado interno)

- (i) La mera exhibición de correlaciones Φ //alguna-variable-de-comportamiento-molar/estado

fenoménico-intencional molar no fija ninguna interpretación unívoca del mecanismo de modulación de la conducta molar, aun en el caso que gracias al empleo de técnicas de fMRI se amplíe el conjunto de variables-de-comportamiento-molar que se puedan correlacionar con Φ ya que la correlación no es causalidad de ningún tipo.

- (ii) El estudio de correlación impone operacionalizar variables proléptico-intencionales propias del vocabulario psicológico, y dicha operacionalización no es en modo alguno autoevidente.
- (iii) Dada su extrema extensión, el concepto de modulación será un concepto de textura abierta que excede la mera noción de causalidad (aún en el caso que esa pudiera definirse en modo preciso).

Pero dado que el Neuromarketing pretende generar tecnologías de intervención e influencia en el comportamiento del consumidor, será necesario resolver estos tres puntos y suplir el análisis cuantitativo- correlacionar con un elemento

teórico que fije de modo preciso el componente de modulación:

Conjunto de elementos que modulan el comportamiento molar = Conjunto de correlaciones psico-físicas + Esquemas de interpretación teórico.

La existencia del problema de un elemento teórico, fue reconocida en forma precisa en un trabajo de fundamentación teórica elaborado por John-Dylan Haynes y Alex Cleeremans (Haynes &.al 1999, 388) donde reconocen que el programa de elaboración de un Correlato Neural de la Conciencia (NCC) ofrece ciertos problemas metodológicos y conceptuales i.e.:

- Problema de medición i.e. “....there are serious measurement problems that stand in the way of correlation research, even after the “hard problem” has

been set aside...” en particular en el caso que intervengan variables cualitativas³; y

- Problema de identificación “... even if one had identified certain internal functional states as indicators of phenomenal states, the empirical psychologist would still be confronted with fundamental problems, such as determining the absence or presence of these functional states.”

Haynes & Cleeremans señalan que dada la magnitud de dichos problemas, la neurofisiología podría emprender la labor de solucionarlos si “...philosophy of science may help and provide **a metatheoretical framework** for the current interdisciplinary project.”.

³ Esta cuestión exactamente la misma que se generó con respecto de la medición de la inteligencia. En contra del optimismo positivista que vio en los test de inteligencia un insumo para una psicología científica y biológicamente fundada; de acuerdo con dichas teorías Inteligencia = Propiedad medida con el test de Matrices Progresivas (o análogo); pero las críticas devastadoras tanto desde la biología –Stephen Jay Gould, Richard Lewontin- como desde la psicología –León Kamin, Howard Gardner- han reducido tal pretensión a Propiedad medida con el test de Matrices Progresivas (o análogo) = cierta propiedad psicológica que está asociada con algún tipo de inteligencia en un momo no determinista estricto difiere categóricamente de la tesis anterior.

Para Haynes & Cleeremans el intento del programa NCC tiende a hallar correlaciones sistemáticas entre el dominio neural y el dominio psicológico –cuya característica central es su carácter proléptico-intencional- empleando técnicas de mapeo cerebral no supone una novedad teórica sino que dicho programa supone retomar el viejo programa psicofísico de G. T. Fechner: donde Fechner usaba mediciones externas de variaciones perceptivas, ahora se hacen mapeos internos en tiempo real. Y dados (i)-(iii), las correlaciones que se hallaren tendrán interpretaciones precisas y estandarizadas con la concurrencia de los ***metatheoretical framework*** que permitirán la operacionalización y traducción del vocabulario proléptico-intencional neurales.

II. La cuestión de los *methateoretical frameworks*.

La investigación de los candidatos a ***methateoretical frameworks*** supone tener una mínima respuesta a la pregunta ¿Cuál es la contribución que puede hacer un ***methateoretical frameworks*** en el desarrollo de una ciencia fáctica? La cuestión no es menor porque una larga tradición filosófica señala que las ciencias fácticas –naturales o sociales- no tienen necesidad de tales ***methateoretical***

frameworks, desde la escuela positivista de A. Comte hasta la crítica radical de estos que hizo Karl Popper⁴. No es este el lugar para ofrecer una defensa articulada de la tesis de la inevitabilidad de los **methateoretical frameworks**; por mor de argumentación asumimos como un supuesto que hay dos modos de considerar a las ciencias empíricas:

- (i) **Ciencia fáctico social & natural** = Parte – sintética- empírico/registrable + Componente – sintético- teórico⁵.
- (ii) **Ciencia fáctico social & natural** = Parte – sintética-empírico/registrable + Componente - sintético-teórico + Componente sintético *a priori*.

⁴ Cf. "The myth of framework can be stated in one sentence... A rational and fruitful discussion is impossible unless the participants share a common framework of basic assumptions or at least, unless they have agreed on such a framework for the purpose of the discussion. This is the myth I am going to criticize" (Popper, 1995, ps. 34-5).

⁵ Este esquema es indiferente a la forma general que tenga el componente teórico i.e. si es un conjunto de axiomas formalizados, o si están expresados en forma intuitiva, o si se expresa por medio de un predicado conjuntista que define familias de aplicaciones y grados de teoriedad.

y que los autores analizados aceptan explícitamente (ii). Como comentario mínimo sobre (ii) se puede señalar que toda teoría empírica supone ciertos principios de clausura disciplinar, y ciertos principios de simplicidad teóricas y de inferencia material que no son datos empírico-registrables ni hipótesis-sobre-mecanismos-o-estructuras etiológicas sino que opera metateóricamente. Y la filosofía de las ciencias ha abordado el análisis de tales elementos metateóricos que operan en el interior de la propia práctica científica como ser entre otros los paradigmas en T. Kuhn, los programas metafísicos de investigación en K. Popper, presuposiciones metafísicas en R. Collingwood y Nicholas Maxwell, los *categorical frameworks* en Stephen Körner, los estilos de pensamiento en L. Flek, los principios guía en Ulises Moulines. Y precisamente esa es la función que Haynes & Cleeremans proponen para el ***methateoretical framework***: generar un metalenguaje que permita la integración de vocabularios/ontologías categorialmente diferentes.

Si NCC exige la consideración del vocabulario proléptico-intencional, los ***methateoretical framework*** que se provean deberán resolver los problemas de medición-operacionalización y de identificación para reinterpretar la

modulación como la concurrencia de ciertas correlaciones entre variables neurales y variables proléptico-fenoménico-intencionales que hayan sido operacionalizadas y de ciertos mecanismos causales por medio de la identificación las clases de eventos correlacionados. Y este es un problema conceptual, pues de los datos registrados no surge un único conjunto de respuestas para los problemas señalados anteriormente; por ello de la presentación que hacen Haynes & Cleeremans se pueden excluir los ***methateoretical frameworks*** estrictamente biológicos –como sería un ***methateoretical framework*** evolucionista- y se debe exigir que incluyan relaciones categoriales entre el vocabulario proléptico-intencional y el vocabulario neuro-fisiológico.

La filosofía de la mente ha discutido tres ***methateoretical frameworks*** que contienen en su núcleo tales relaciones categóricas:

(ME) Materialismo eliminativo: Esta posición radical ha sido definida por Paul Churchland: “Eliminative materialism is the thesis that our commonsense conception of psychological phenomena constitutes a radically false theory, a theory so

fundamentally defective that both the principles and the ontology of that theory will eventually be displaced, rather than smoothly reduced, by completed neuroscience.” (Churchland, Paul (1981), 67). No hay correspondencia entre ambos léxicos, ya que no hay equivalencia extensional entre sus categorías y por consiguiente no hay ningún conjunto de verdades comunes ni hay ninguna convivencia entre el vocabulario proléptico-intencional⁶ y el vocabulario

⁶ En este trabajo se usará la expresión *vocabulario/entidades proléptico-intencionales*, en vez de la tradicional caracterización *intencional*, dado que esta última expresión es ambigua. En efecto, desde el desarrollo de la fenomenología y la psicología comprensiva la noción de intención se emplea con el sentido técnico que le dio Brentano i.e. “Todo fenómeno psíquico está caracterizado por lo que los escolásticos de la Edad Media han llamado la inexistencia <existencia-en> intencional (o mental) de un objeto, y que nosotros llamaríamos, si bien con expresiones no enteramente inequívocas, la referencia a un contenido, la dirección hacia un objeto (por el cual no hay que entender aquí una realidad), o la objetividad inmanente. Todo fenómeno psíquico contiene en sí algo como objeto, aunque no siempre del mismo modo. En la presentación hay algo que es presentado; en el juicio algo viene aceptado o rechazado; en el amor, amado; en el odio, odiado; en el deseo, deseado, etc.”. (Brentano, Franz 1935/ 1874, 81-2). Esta definición técnico-filosófica de *intención* difiere del uso mundano de *intención* como “...tener la idea de realizar tal y cual cosa...”, o “...tener el propósito de lograr tal resultado...”, sin embargo el sentido mundano de intención es el que aparece en el vocabulario del economista, ya que los estados intencionales que primariamente le interesan son los estados propositivos i.e. elegir un satisfactor para una preferencia en función de su valuación y de la utilidad esperada o imaginada. Dado que el significado mundano de *intención* es parte del significado técnico (en el sentido que tener la intención de hacer tal y cual cosa es él mismo un estado intencional) llamaremos intencionalidad

neural; este modelo da lugar a la Tesis de la paráfrasis de las oraciones proléptico-intencionales. Pero los teóricos eliminativistas aceptan que el vocabulario proléptico-intencional puede tener un rol comunicacional preciso, y no renunciabile.

(TI) Materialismo estricto o teoría de la identidad: La teoría de la Identidad fue propuesta como la concepción materialista más intuitiva, y su formulación estándar fue hecha por U.T. Place y H. Feigl. U. T. Place sostenía que “I am no clamming that statements about sensations and mental images are reducible to or analyzable into statements about brain processes...I assert... is that the statements consciousness is a state of the brain... is a reasonable scientific hypothesis <y> a general or universal proposition applying to all states of consciousness whatever <y lo que se aserta es una identidad de composicion que ocurre en aquellos casos que> the technical scientific observation sets in the context of the appropriate body of scientific theory provide an immediate

propia al significado proléptico o propositivo y llamaremos intencionalidad impropia al sentido mundano de intención.

explanation of the observation made by the <observador común>” (Place, (1956), 46-8). De acuerdo con Place no hay un reemplazo lexical, pues cada léxico mantiene independencia lógica, si bien el lenguaje mundano de atribución de estados intencionales y actitudes proposicionales debe identificarse-por-composición-tipo-por-tipo con el lenguaje neurofisiológico → Tesis de la identificación de los léxicos.

Si bien el materialismo eliminativo y la teoría de la identidad difieren sobre la supervivencia o no del vocabulario proléptico-intencional, ambas posiciones afirman la primacía epistemológico/ontológica del vocabulario y las entidades neurofisiológicas sobre las proléptico-intencionales; si tomamos el ejemplo anterior, la teoría del consumidor reemplazará su viejo ropaje proléptico-intencional por otro neurofisiológico, el consumidor dejará de tener deseos, actitudes y creencias, y paulatinamente pasará a tener circuitos reflejos, estimulaciones de terminales, asambleas neuronales, y la tendencia será firme y irreducible a considerar al consumidor como un conjunto de sistemas neuro-fisiológicos, sometidos a estímulos físicos-químicos desde el entorno. Tomando el estudio anterior ¿podríamos

decir que Φ es la causa del gusto por Coca-Cola, y en ese caso que las campañas de Coca-Cola deberían emplear señales, colores y formas que estimularan Φ ? O ¿debería contener Coca-Cola alguna sustancia que altere la corteza dorsolateral prefrontal?

Diferencias de lado, ME y TI sostiene una visión metateórica *determinista-de-manipulación*⁷ y químico-biológica de la agencia humana. Pero ¿es ello así? ¿Puede sostenerse que la investigación de la agencia humana puede desarrollarse de este modo? Esta tesis solo puede afirmarse si fuera posible mostrar que el léxico/ontología proléptico-intencional es completamente dispensable, aún en el terreno de falsedades útiles y comprensibles; de lo contrario será necesario indagar un *methateoretial framework* que cumpla las tareas que piden Haynes & Cleeremans y que no tenga esta consecuencia.

⁷ Es necesario distinguir el problema metafísico del determinismo, del problema técnico-científico de la intervención determinista. Es posible sostener que el comportamiento que se supone libre no lo es en absoluto, porque hay una serie de determinantes, pero podría ser el caso que no haya ninguno de estos determinantes operacionalizados para intervenir y dirigir el comportamiento del agente en una dirección prefijada y con un grado de éxito predecible; estas serán las tecnologías de manipulación de la conducta o determinismo de manipulación.

III. Indispensabilidad del vocabulario proléptico-intencional.

Históricamente el estudio de la decisión fue influenciado por matemáticos y economistas interesados en las propiedades formales de la decisión considerada como elección entre *inputs* de información con diversos grados de limitación, y ejecutadas por un agente abstracto caracterizado por parámetros definidos⁸, y en ese caso el interés fue puesto en las formas de representación, evaluación, comparación y cálculo de dicha información y en el establecimiento de criterios normativos de eficacia y excelencia en dicha

8 Se trata del Modelo del Homo oeconomicus, cuya caracterización precisa hiciera en un modo preciso John Stuart Mill v.g. "What is now commonly understood by the term "Political Economy"... is a branch of that science <la Economía Política examina la acción humana>...solely as a being who desires to possess wealth, and who is capable of judging of the comparative efficacy of means for obtaining that end. It predicts only such of the phenomena of the social state as take place in consequence of the pursuit of wealth. It makes entire abstraction of every other human passion or motive; except those which may be regarded as perpetually antagonizing principles to the desire of wealth, namely, aversion to labour, and desire of the present enjoyment of costly indulgences. These it takes, to a certain extent, into its calculations, because these do not merely like other desires, occasionally conflict with the pursuit of wealth, but accompany it always as a drag, or impediment, and are therefore inseparably mixed up in the consideration of it." (Mill, 1844: 137-8).

elección. Sin embargo el desarrollo de la experimentación en ciencias sociales –tanto psicología como economía y ciencias políticas- ha aportado una dimensión diferente y se ha hecho necesario vincular las predicciones modelo-teoréticas desde el *homo oeconomicus* con los resultados tanto de diseños experimentales como de estudios cuasi-experimentales, los que proveyeron abundante evidencia que muchas veces contravino dichas predicciones y señaló la presencia de otros elementos en la decisión⁹, en la medida que dichos factores son factores intensionales suponen contextos de habla; actitudes proposicionales, reglas de inferencia material y escenarios perceptuales que influyen en los *inputs* y en las elecciones en formas que no se pueden predecir; pues tales factores intensionales suponen una terminología y una ontología proléptico-intencional. Dichas terminología y ontología proléptico-intencional son inevitables en el estudio de la decisión humana como forma específica

⁹ Los experimentos sobre decisión en contextos de riesgo y de evaluación condicional y de interpretación de informes en los, mostró el carácter dependiente de descripción de los objetos de preferencia (*framing effects*), la no neutralidad del punto de partida (*starting point*) y la primacía de las decisiones por defecto (*default rule*).

de los agentes sociales de efectuar elecciones reales y es irreductible a meras secuencias de estados-eventos neural

Un argumento que abona dicho carácter irreductible del vocabulario/ontología proléptico-intencional es el argumento que ofrece Donald Davidson a favor del monismo anómalo, argumento que apela al holismo intrínseco de dicho vocabulario/ontología; si bien este argumento se propuso en el marco de una teoría metafísica de lo mental (argumento holista fuerte) puede entenderse con independencia de su carácter metafísico, en el marco de la representación en una teoría psicológica y una ontología informática del vocabulario proléptico-intencional (argumento holista débil). Aún si se rechazan concepciones metafísicas constructivistas o sistémicas, el holismo semántico supone la intraducibilidad tipo a tipo entre tipos de diferentes léxicos categoriales, dado que la determinación de cada tipo en un léxico categorial será una función de una trama local –que variará de léxico categorial a léxico categorial.

Este argumento apunta a la incorporación de supuestos de carácter sistémico y estructural, al menos en el sentido del holismo débil en la teoría de la decisión. Esta imbricación

entre teoría de la interpretación –que da cuenta de los estados proléptico-intencionales- y teoría de la decisión ha sido señalada por Donald Davidson “La teoría de la interpretación verbal y la teoría de la decisión bayesiana están evidentemente hechas la una para la otra. La teoría de la decisión debe liberarse de la asunción del acceso independiente de los significados; la teoría del significado necesita una teoría del grado de creencia para hacer un uso serio de las relaciones de apoyo evidencial. Pero afirmar estas dependencias mutuas no es suficiente, porque ninguna teoría puede desarrollarse primero como una base para la otra. No hay ninguna forma de añadir simplemente una a la otra porque cada una para empezar requiere un elemento extraído de la otra. Lo que se busca es una teoría unificada que ofrezca el grado de creencia, las deseabilidades en una escala de intervalos, y una interpretación del habla, *una teoría que no asuma que los deseos o las creencias tienen que individualizarse de antemano, mucho menos que se cuantifiquen.*” (Davidson, Donald (1997/1990), 197).

En este sentido el vocabulario categorial teórico-modelístico de las ciencias económicas posee un sub-léxico comportamental que podemos llamar –en analogía con la

filosofía de la mente- ***folk psychology o psicología mundana del economista (PME)***. En la medida que el vocabulario proléptico aparece involucrado directamente la teoría de la decisión cualquier caracterización de los procesos de elección y decisión, será imposible una irreductibilidad tipo a tipo entre el léxico/ontología de PME y el modelo de vinculación mecánica de *inputs* y preferencias del ***homo oeconomicus***.

Un ejemplo –ya clásico- permite ilustrar esto apelando al capítulo VII “On Intentionality” de los *Principles* de J. Bentham; aquí la noción de *intentionality*, se toma con su significado impropio, pues se refiere al comportamiento proléptico.

En ese lugar se caracteriza a quien toma decisions como un actor, cuyos actos debieran ser al menos “...intentional in the first stage, it is no act of yours: there is accordingly no intention on your part to produce the consequences...” ya que imputar propositividad a las consecuencias supone que: “... the consequences of an act cannot <sean directas u oblicuas> be intentional, without the act’s being itself intentional in at least the first, stage...”. (Bentham 70).

La imputación de propositividad –y la forma que ella tenga- se hace a los actos, en tanto episodios conductuales que se conectan etiológicamente con la voluntad, pero no a las consecuencias o efectos que se conectan causalmente con el curso de acción i.e. “A man intends the act: and by his intention produces the act: but as to the circumstances, he does not intend *them*: he does not, inasmuch as they are circumstances of it, produce them. If by accident there be a few which he has been instrumental in producing, it has been by former intentions, directed to former acts, productive of those circumstances as the consequences: at the time in question he takes them as he finds them. Acts, with their consequences, are **objects of the will** as well as of the understanding: circumstances, as such, are objects of the understanding only”. (Bentham, 76, destacado nuestro). En este esquema brevemente se puede señalar que los tipos del léxico de propositividad son categorialmente diferentes a los tipos de la trama causal que está lejos de la propositividad del agente.

Los *Principles* reposan sobre estas caracterizaciones propositivas y a partir de allí ofrecen una clasificación exhaustiva de los tipos conductuales, analizando cada

parámetro constitutivo, sus grados y combinaciones, y en cada resultado analiza la situación respecto del principio de utilidad y el criterio de placer y dolor. Pero cada paso que ofrece mantiene como vocabulario primitivo el elemento proléptico (*to intend* y *objects of the will*), lo que entraña que el vocabulario utilitarista es en forma inescapable impropriamente intencional y por ello intencional *tout court*.

La indispensabilidad de la intencionalidad impropia –y de la intencionalidad *tout court*- impone ciertos límites en la elección de un ***methateoretical framework***; en efecto, si bien es posible discrepar sobre el quantum de intencionalidad impropia que suponen las teorías económicas actuales; la asunción de ME o TI supone limitaciones importantes en la caracterización precisa de dicho componente en la medida que ME o TI suponen vectores de investigación y de interpretación de naturaleza neurológica.

IV. Un examen del funcionalismo homuncular.

Dada la irreductibilidad relativa del léxico/ontología proléptico-intencional se hace necesario examinar aquellos ***methateoretical frameworks*** que permitan la integración del vocabulario proléptico-intencional y del vocabulario neural en forma no reductiva y no dispensable; ello nos lleva a examinar el tercer ***methateoretical framework*** seleccionado: el funcionalismo de dos niveles.

Funcionalismo: El funcionalismo tiene su carta de nacimiento es el célebre artículo de Alan Turing que sugirió definir los tipos psicológicos –como *inteligencia, resolución de problemas, decisión racional* y demás- en términos de reglas de enlace entre inputs y outputs de un sistema que responde consignas verbales. De ese modo el funcionalismo introdujo en la filosofía de la psicología la clásica analogía del ordenador i.e. “The seductive comparison of people (or their brains) to computing machines drew our attention to the contrast between a machine's program (abstractly viewed) and the particular stuff of which the machine happens to be physically made, that realizes the program” (Lycan, 1995, 36). Esa corriente se desarrolló a partir de trabajos de J. Searle y H. Putnam hasta convertirse en la filosofía de la mente dominante en los 70 y 80, en particular porque opero como

methateoretical frameworks del desarrollo de las ciencias cognitivas y los enfoques cognitivos en ciencias sociales y filosofía. Según Putnam el programa "...try to show that all of the issues ... and puzzles that make up the traditional mind-body problem... arise in connection with any computing system capable of answering questions about its own structure, and have thus nothing to do with the unique (if it is unique) of human subjective experience" (Putnam, 1975/1966, 362). Si el sistema autorreferente –cuyo modelo será la máquina de Turing- se caracteriza sus estados internos de cierto modo, entonces el conjunto de descripciones de los estados de dicho sistema al proveer tal caracterización, será todo lo necesario para identificar dicho estado interno, y en diferencia con la singularidad que se pretendía de los estados mentales subjetivos, podrían tener realizaciones diversas, siempre que se satisfagan las descripciones de estados internos. En esta concepción el léxico proléptico-intencional se definirá por medio de conjuntos de descripciones de estado característicos y sin referencia a la naturaleza y materiales del sistema que toma dichos estados. En consecuencia F. puede entenderse como la diferencia entre rol/realización y ocupante/estofa. W. Lycan caracteriza esta distinción "...between role and

occupant...” como central y sostiene “... It is the former, not the latter, that interests us *vis-a-vis* the interpretation, explanation , prediction, and exploitation of the machine's "behavior What matters is function, not functionary; program, not realizing -stuff; software, not hardware; role, not occupant”. (Lycan, idem). De ese modo una teoría funcionalista determinará el conjunto de roles deterministas – estrictos o probabilistas- que determinan cada tipo de estado mental, v.g. tener dolor, decidir, elegir, predecir, evaluar riesgos, etc. → Tesis de la autonomía de los léxicos.

Así presentado el funcionalismo parecería proveer de un ***methateoretical frameworks*** que acepta la pertinencia lenguaje proléptico-intencional de estados subjetivos y actitudes proposicionales. Pero no parece intuitiva la traslación de la dicotomía rol-ocupante a la dicotomía cerebro-conducta, pues la primera parece ser una dicotomía complementaria mientras que la segunda parece no ser una dicotomía en absoluto, ya que para decir que el cerebro actúa se suponen una serie de mediaciones tales como *cuerpo propio, aparato psíquico, subjetividad, red vincular, instituciones sociales*, etc. hasta llegar efectivamente a la noción de agente. Pero si este fuera el caso, no quedaría

claro el rol de las neurociencias en ese esquema, dado que el cerebro literalmente no actúa; bajo este enfoque lo único que proveerían las neurociencias sería un sistema de transliteración absolutamente grosero –al menos tan grosero como la transliteración de comportamiento molar y vincular en términos genéticos, como creyeron algunos teóricos reduccionistas que suponían haber hallado el gen del amor u otros- que se limitaría a proveer de analogías neurales para el comportamiento, con las cuales no podrían caracterizarse los estados subjetivos o vinculares.

Pero si este desacople o distancia entre ambos pares conceptuales parece un problema para el funcionalismo a secas, ella puede abordarse por medio de una variedad del funcionalismo: el llamado **funcionalismo homuncular** como diferente al **funcionalismo de dos niveles** o **funcionalismo simpliciter**.

La teoría del funcionalismo homuncular fue propuesta por Daniel Dennett y William Lycan como una respuesta a ciertos argumentos que atacaban al funcionalismo de dos lados. Uno de esos argumentos es el llamado **argumento de la mente China o de la realización arbitraria**, propuesto por

Ned Block (cf. Block, 1993). Este argumento era un experimento mental que Block definía del siguiente modo:

Suppose we convert the government of China to functionalism, and we convince its officials to realize a human mind for an hour. We provide each of the billion people in China (I chose China because it has a billion inhabitants) with a specially designed two-way radio that connects them in the appropriate way to other persons and to the artificial body mentioned in the previous example. We replace each of the little men with a citizen of China plus his radio. Instead of a bulletin board we arrange to have letters displayed on a series of satellites placed so that they can be seen from anywhere in China...The system of a billion people communicating with one another plus satellites plays the role of an external "brain" connected to the artificial body by radio... It could be functionally equivalent to you for a short time, say an hour. I count as inputs and outputs only neural activity in the artificial body connected by radio to the people of China."
(Block;(1980/1993) 239)

En este sistema estarían presentes todos los elementos para una caracterización funcional de un estado mental i.e. *inputs*

sensoriales, *outputs* conductuales, y relaciones causales entre estados internos. De ese modo según la doctrina estándar del funcionalismo cada estado total de la sociedad china en un momento podrá ser descrito por una serie de tablas de estados funcional que caracterizarían estados mentales, v.g. una sensación de dolor o placer. Pero – concluye Block, parece completamente contraintuitivo atribuir mente y conciencia al sistema material que conforma la totalidad de los chinos y las señales de radio que intercambian entre si, y en suma al sistema Sociedad china +cuerpo artificial, en general estos argumentos reposan sobre la insuficiencia de criterios para atribuir estados intencionales o concientes a sistemas materiales no subjetivos.

Estos argumentos parecían aplicarse bien al funcionalismo de dos niveles; por ello Dennett y Lycan exploraron la posibilidad de un funcionalismo de niveles múltiples. En esa dirección W. Lycan ofreció una versión de la teoría funcionalista que preservaba la distinción rol-ocupante pero rechazaba la tesis de los dos lados i.e. "Very generally put , my objection is that software "/"hardware " talk encourages the idea of a bipartite Nature , divided into two levels, roughly the physiochemical and the (supervenient) " functional " or higher -organizational

- as against reality, which is a multiple hierarchy of levels of nature , each level marked by nexus of nomic generalizations and supervenient on all those levels below it on the continuum “ (Lycan, 1995, 37), de ese modo a la dicotomía rol-ocupante le correspondería una multiplicidad de estados cerebro – X_1 – X_2 - X_3 ... conducta.

El funcionalismo homuncular puede verse como una metateoría cuyas tesis básicas son las siguientes:

FH1: Tesis de la jerarquía múltiple: la tesis básica del funcionalismo homuncular es la tesis de los niveles de la naturaleza; el carácter funcionalista surge cuando esta jerarquía no se analiza de modo mereológico¹⁰ –i.e. a partir de relaciones de partitudo- sino funcional como define el autor “...the " function " /" structure " distinction goes relative: something is a role as opposed to an occupant, a functional

¹⁰ Una jerarquía mereológica supone que los elementos de un nivel inferior están compuestos por partes constituyentes de entidades que serán elementos del nivel superior i.e. los niveles de integración en biología.

state as opposed to a realizer, or vice versa, only modulo a designated level of nature” (Lycan, ídem). Estos niveles no solo tienen integración vertical, la que es propia de la jerarquización mereológica, sino integración horizontal; como señala Lycan una célula está compuesta por núcleo y membrana, que a su vez tienen entre si relaciones funcionales, sin ser partes uno de otra.

FH2. Tesis homuncular¹¹: Es la tesis constitutiva de FH, y establece que la tarea explicativa del comportamiento molar de un agente personal debe hacerse apelando a una escala de subniveles que se desagregan unos de otros: “To

¹¹ De origen alquímico –cf. Geber (Abu Musa J 721–815) y Paracelsus (Philippus Aureolus Bombast von Hohenheim 1493–1541) como sus fuentes directas- el concepto de homúnculo (del latín *hombre pequeño*) fue un concepto básico de la embriología del s. XVII, introducido para explicar el desarrollo embriológico humano, a partir de la presencia en los espermatozoides de un humano muy pequeño que luego crecía hasta llegar a la criatura alumbrada (incluso se ofrecieron dibujos de tales homúnculos). En general una teoría es homuncular cuando explica un cierto ítem postulado que existe alguna versión pre-formada de X, o un X homuncular. No necesariamente las teorías homunculares son teorías que cometen la falacia del círculo vicioso.

characterize the psychologists quest in the way I have is to see them as first noting some intentionally or otherwise psychologically characterized abilities of the human subject at the level of data or phenomena , and positing - as theoretical entities - the homunculi or sub- personal agencies that are needed to explain the subject 's having those abilities... Then the psychologists posit further, smaller homunculi in order to explain the previously posited molar behavior of the original homunculi, etc., etc.” (Lycan, 1995, 40).

La tesis es extremadamente sencilla en su concepción: para explicar un rasgo proléptico-intencional central de la agencia personal, se postulan niveles explicativos de agencia sub-personal que posean propiedades intencionales; pero lejos de caer en circularidad –como sería el caso de las explicaciones homunculares clásicas- esta versión introduce la tesis de la jerarquía múltiple, o se que se postula una intencionalidad en grados decrecientes y cada vez más difusa, vinculadas por relaciones funcionales, hasta llegar a niveles jerárquicos cuyo grado de intencionalidad es cero y que se pueden caracterizar en términos de leyes causales neurales: “ We explain the successful activity of one homunculus , not by idly positing a second homunculus within it that successfully

performs that activity, but by positing a team consisting of several smaller, individually less talented and more specialized homunculi - and detailing the ways in which the team members cooperate in order to produce their joint or corporate output . (Lycan, 1995, 40).

FH3. Tesis de la sinonimia transistémica: Esta tesis funge como parte de una tesis ontosemántica: “I propose to type-identify a mental state with the property or having such-and-such an institutionally characterized state of affairs obtaining in one (or more) of one’s appropriate homuncuional departments or subagencies.” (Lycan, 1995, 41). Esta tesis permite recoger la clásica intuición, defendida entre otros por D. Davidson y W. Sellars, de la irreducibilidad de los tipos psicológico-institucionales a tipos fisiológicos de cualquier nivel de descripción, ya que “The apparent irreducibility of the mental is the genuine irreducibility of institutional types to the less teleological” (Lycan, 1995, 42). Mientras en Davidson el holismo semántico opera sobre todo el vocabulario psicológico y fisiológico, la tesis de Lycan parece desagregarlo: cada tipo institucional que corresponde a un

tipo psicológico, es un sistema descendente de vínculos funcionales organizados –que él ilustra con el diagrama de flujo desde un nodo hacia los haces descendentes y paralelos. Es el complejo de realizaciones el que opera todo junto, y esta condición holista no es necesaria en el lenguaje categórico del neurofisiólogo. En ese sentido el vocabulario psicológico no será reemplazado por un vocabulario de neurólogo sino por un vocabulario de niveles y formas de integración.

FH4. Tesis teleológica: Si los niveles de integración se definen por relaciones funcionales o roles que se ejecutan en materialidades biológicas de diversa estructura, toda la red y los árboles de integración ascendente-descendente están vinculados de una u otra manera por relaciones biológicas. Y así como el componente homuncular finca niveles de intencionalidad descendente en una secuencia de-arriba-a-abajo en el nivel molecular, el componente teleológico finca niveles de determinación biológica decreciente en la secuencia de-abajo-a-arriba.

Lycan identifica el componente teleológico no con su variante escolástica¹² sino con la versión evolutiva “I hope, and am inclined to believe, that the teleological characterizations that Homunctionalism requires can be independently explained in evolutionary terms. This hope is considerably encouraged by the work of Karl Popper, William Wimsatt, Larry Wright, Karen Neander, and other philosophers of biology; I cannot improve on their technical discussions.” Esta tesis general puede aplicarse al enfoque homuncular: “... the teleologicalness of characterizations is a matter of degree: some characterizations of a thing are more teleological than others...<por ello, la existencia de niveles de descripción asociados a jerarquías homunculares hace que> we cannot split our theory of nature neatly into a well-behaved, purely mechanistic part and dubious, messy vitalistic part better ignored or done away with. And for this reason we cannot maintain that a reduction of the mental to the teleological is no gain in ontological tractability; highly teleological characterizations, unlike naive and explicated mental

¹² En la filosofía natural escolástica la teleología se vinculaba con la existencia de causas finales, y especialmente Dios como Causa final Perfecta.

characterizations, have the virtue of shading off fairly smoothly into (more) brutally physical ones “

El análisis de un ejemplo -la capacidad de reconocimiento de rostros postulando una macro-estructura funcional interna concebida para dicho fin y desagregada en niveles funcionales homunculares- permitirá reconocer que “...the institutional hierarchy (from recognizer to scanner to light meter to photosensitive substance and as much further down as one might care to go) <no tiene un punto fijo de transición en el que >... stop being teleological, period, and start being purely mechanical, period <de modo general que>... there is no such point , but rather a finely grained continuum connecting the abstract and highly teleological to the grittily concrete and only barely teleological” (Lycan, 1995, 44)

Es esta organización funcional homuncular la que le hace que consideremos al vocabulario y la realidad psicológica como totalmente distinta de la neuro-fisiológica, sin que metafísicamente lo sean.

FH5. Tesis de la continuidad normativa: Si bien esta tesis no forma parte de la propia presentación de Lycan –en particular porque su preocupación es la filosofía de la mente-, podemos señalar que por un lado es consistente con FH, pero más importante aún es necesaria para dar contenido al nivel de agencia transpersonal que no le interesa tanto al autor, pero que es relevante en la filosofía de la economía. Así como de acuerdo con FH1-FH4 el vocabulario categorial de la economía será en algún nivel irreductiblemente teleológico-funcional, en otros niveles será homuncular + neural y en algún nivel será simplemente neural, podemos decir que ese vocabulario tendrá algún nivel descriptivo + normativo, y que el mismo elemento teleológico que opera desde los niveles biológicos más elementales, se podrá extender sin problemas desde los niveles de agencia subpersonales hasta aquellos completamente suprapersonales. Es factible conectar esta tesis con los desarrollos de los economistas institucionales v.g. John Commons quién define la acción colectiva v.g. “Collective action ranges all the way from unorganized custom to the many organized going concerns, such as the family, the corporation, the trade association, the trade union, the reserve system, the state. The principle common to all of them is

greater or less control, liberation and expansion of individual action by collective action <de modo que> ... the collective control takes the form of a taboo or prohibition of certain acts, such as acts of interference, infringement, trespass; and this prohibition creates an economic status of liberty for the person thus made immune.” (Commons, 1931, 249). La tesis de Commos preanuncia las tesis constructivistas de la metafísica social de autores como J. Searle, A. Schutz o H. Putnam; aquí no interesa tanto su linaje y alcance, sino señalar su pertinencia teórica: lejos del dogma básico de la filosofía analítica que afirmaba en forma absoluta la cesura entre el ser y el deber ser, esta tesis integra en el mismo enfoque teleológico-funcional las normas e instituciones centrales de la agencia trans-personal.

En resumen el enfoque **funcional-homuncular** no culmina en los organismos corpóreos, sino que las jerarquías ontológicas se extienden más allá de la dimensión personal, hasta entidades colectivas integradas en el sistema de niveles jerárquicos. En ese sentido, suponiendo la identificación entre **agente = organismo corpóreo humano = persona**, este cuadro resume los tres niveles marco de agencia que reconoce FH:

<p>Nivel III: Supra-agencial o agencia colectiva: entidades colectivas compuestas teleológicamente por sujetos individuales corpóreos</p>	<p>“Organisms, for that matter, collect themselves into organized (<i>organ-ized</i>) groups. And there is no clear difference of kind between what we ordinarily think of as single organisms and groups of organisms that function corporately in a markedly single-minded way- “group organisms” themselves, we might say...” (Lycan, 1985, 38).</p>
<p>Nivel II: Agencial o agencia personal: sujetos individuales corpóreos compuestos teleológicamente por constituyentes sub- agenciales</p>	<p>“...a person is a corporate entity that corporately performs many immensely complex functions... “ (Lycan, 1995, 40)</p>

<p>Nivel I Sub-agencial o agencia funcional: niveles constituyentes de los individuos corpóreos</p>	<p>“...this person <es descripta> by means of a flow chart which depicts the person’s immediately sub-personal agencies and their many and various routes of access to each other...” (Lycan, 1995, 40)</p>
--	---

Estos son niveles marco y cada uno de ellos contiene subniveles que corresponden a algún nivel de abstracción; además dichos niveles no solo actúan en jerarquía de niveles, sino que operan tanto lateral como transversalmente; del mismo modo no hay límites o transiciones absolutas entre un sub-nivel y otro; la Tesis Teleológica permite pensar transiciones de fase subjerarquizadas de cualquier orden de magnitud que sea necesario.

Si el científico del comportamiento se interesa primariamente por los niveles marco I y II y sus desagregados, el caso del economista es diferente: sistemáticamente se ocupa de los niveles II y III y sus desagregados, y solo en los últimos años

apareció un vector de explicación hacia el nivel I. Este vector supone la introducción de niveles de agencia sub-personal que provienen de las neurociencias. Dicho de otra manera, más allá del teleologismo entendido en términos de una componente de vectores evolutivos que operan sobre rasgos genotípicos, fenotípicos y comportamentales, podemos introducir sin violencia con este desarrollo una deriva normativa-institucional; y en ese sentido podemos extender la misma ontología de integración de niveles desde lo puramente mecánico-etiológico químico hasta lo puramente intencional-teleológico-normativo, más allá de los intereses del filósofo de la psicología hasta los intereses de la filosofía de la economía.

V Alegato por FH

La asunción de FH como el *metatheoretical framework* que opere en el encuentro de las neurociencias y ciencias económico-sociales, supone “dar vuelta la taba” en el sentido que plantea la comunicación –vía los niveles homunculares-

del léxico/ontología proléptico-intencional y las neurociencias en el proyecto de una ciencia de la elección en contextos de escasez, sesgos limitativos y condiciones contextuales.

El objetivo del economista es la explicación y predicción de las acciones que ciertos agentes ejecutan para procurarse ciertos satisfactores que son escasos, que tienen aplicaciones alternativas, cuyas tasas de escasez relativa es una función de las acciones de terceros agentes y cuya realización depende de contextos de información limitada y tasas variables de riesgo o incertidumbre, así como las propiedades de los agregados de satisfactores y las interacciones que dichos agregados tengan con las trayectorias de comportamiento individual.

Esta tarea es llevada a cabo desde teorías con principios y supuestos de diverso nivel de abstracción, y por medio de modelos fundados en dichos supuestos en los que se generan ciertas predicciones-en-el-modelo. El proceso de contrastación empírica ha señalado que ciertos supuestos considerados estándares en las teorías económicas admitidas, deben ser revisados o cambiados en forma drástica –en particular supuestos sobre la racionalidad y las

limitaciones de las capacidades de decisión en contextos que tienen los agentes reales. Ello ha llevado en diversas direcciones, una de las cuales es la apelación a los resultados de las neurociencias.

Benoit Hardy-Vallé (Hardy-Vallée, 2007) ofrece un ejemplo de integración de resultados: la vinculación de resultados de neurociencias con ciertos modelos de elección según valuaciones. Al respecto sostiene que “From an internal perspective, neuroeconomics studies show how externally irrational choices can be internally rational. Loss-aversion and ambiguity-aversion are not optimal under decision theory, but neuroeconomics shows that loss- and ambiguity-averse subjects are rational in the internal sense, because they try to minimize negative feeling (elicited by amygdale activation). In this sense, it can be rational to maximize experienced utility, based on predicted utility. Loss-aversion can also be internally rational because preferring a sure decision is a way to minimize the painful feeling of loss and the post-decision feeling of regret that unfair offers trigger, in the responders’ brain, a ‘moral disgust’: the anterior insula (associated with negative emotional states like disgust or anger) is more active when unfair offers are proposed... . Various studies indicate

that a loss of money can elicit negative emotional response (Delgado, Labouliere and Phelps), and that avoiding an aversive event can generate rewarding signals in the orbitofrontal cortex (Kim, Shimojo and Doherty).” (Hardy-Vallée, 2007,944, 947).

En un plano descriptivo y fenoménico, Hardy-Vallée sostiene que hay al menos dos niveles de los que se puede predicar “... es racional”:

(1) Nivel sub-agencial: (racionalidad internista)

En su análisis señala que los agentes tendrían una “...rational in the internal sense...”.

(2) Nivel agencial: “(racionalidad externista)·En su análisis señala que en ese nivel los agentes llevaron a cabo”...externally irrational choices...”.

La interpretación de esta descripción, varía completamente desde el *metatheoretical framework* que se seleccione. En efecto si interpretamos desde TI o ME la cuestión se complica; ME rechaza la relevancia de los predicados proléptico-intencionales y sostiene que dicho vocabulario es literalmente vacuo, en ese caso el nivel de análisis agencial

deberá ser abandonado y la teoría de la decisión, deberá basarse completamente en la racionalidad internista, y en las homeostasis neurales –y en ese sentido habría que desarrollar experimentos económicos con análisis cerebral con técnicas de captura, y luego ir a la base neuroquímica específica. Por otro lado TI sostiene que debe haber una co-extensionalidad entre el vocabulario proléptico-intencional y el vocabulario neural; pero si esto fuera así no se podría predicar racionalidad en un nivel e irracionalidad en el otro sin contradicción lógica.

Por el contrario la interpretación desde FH permite superar estos problemas; de acuerdo con **FH3** las relaciones semánticas del léxico del vocabulario proléptico-intencional y del vocabulario neural son entre redes de relaciones funcionales ascendentes y descendentes o estados institucionales de cosas. En este sentido es perfectamente posible que un agente esté en un estado agencial irracional, que contenga como *input* homuncular en algún nivel jerárquico inferior estados sub-agenciales racionales: precisamente la investigación empírica mostrará que para una categoría de *ser racional* hay entre los dos estados una dependencia etiológica: así el estado emocional asociado con

la aversión al riesgo sería el ocupante de ciertos estados funcionales que tendrán como antecedentes homunculares de nivel sub-agencial ciertos sistemas químico-biológicos de la ínsula anterior; y al mismo tiempo este sistema no es meramente de dos lados porque la complejidad de las respuestas afectivas supone más sistemas que la ínsula anterior, vg. sistemas conectados con el aprendizaje, la autovaloración, la capacidad de administrar crisis, el temple, cada uno de los cuales tendrá antecedentes homunculares, que al mismo tiempo tendrán relaciones horizontales entre ellos¹³.

En ese sentido FH parece mucho más interesante que las propuestas de ME o de TI. Porque al proponer un marco general de interpretación de los resultados neurofisiológicos a partir de las tesis homuncular y teleológica, la relación entre la teoría de la decisión fincada en el nivel de la agencia personal y las teorías que postulan niveles de agencia y

¹³ Pero hay más. Del mismo modo que se integro la racionalidad sub-agencial y la racionalidad-agencial FH permitiría sostener sin contradicción que una racionalidad supra-agencial pueda tener como antecedente homuncular estados agenciales irracionales. Sobre este tema volveremos en un trabajo posterior.

decisión sub-personales, será una relación de integración; y ello llegará incluso a los niveles más comunes al economista desde el agente elector personal hasta el mercado y las determinaciones macroeconómicas, i.e. desde los subniveles más elevados del Nivel II, hasta el nivel III.

VI. Casos de aplicación de FH..

El caso que presenta Hardy-Vallée expone el núcleo de problema de las variables neurales y sirve de ejemplo modelo que permite señalar la pertinencia de FH, pero podemos preguntarnos si este *metatheoretical framework* se puede generalizar en un sentido de hacer relevantes las variables neurales para la comprensión del comportamiento económico. En ese sentido ofreceremos algunos desarrollos en neurociencias, examinaremos sus consecuencias para la teoría económica, y la pertinencia de FH para el caso.

(A) La existencia de un *delay* neural-intencional y el problema del determinismo neural. La existencia de un *delay*

temporal entre la dimensión neural y la dimensión proléptico-intencional fue puesta en evidencia por los experimentos de Benjamin Libet¹⁴; pero el empleo de técnicas de fMRI permitió que un grupo de investigadores del Max Planck Institute for Human Cognitive and Brain Sciences descubrieran un notable incremento en la magnitud de dicho *delay* al observar que "...the outcome of a decision can be encoded in brain activity of prefrontal and parietal cortex up to 10 s before it enters awareness." (John-Dylan Haynes & al. 2008)...En sus experimentos pudieron establecer que había una instancia sub-personal en un proceso de toma de decisiones, y concluyeron que "Our study shows that decisions are

¹⁴ Cf. Libet, B. (1985) 'Unconscious cerebral initiative and the role of conscious will in voluntary action', *Behavioral and Brain Sciences*, **8**, pp. 529–66. En sus experimentos Libet partió de lo que podemos llamar una hipótesis convencional i.e. "In the traditional view of conscious will and free will, one would expect conscious will to appear before, or at the onset, of the RP, and thus command the brain to perform the intended act." (Libet (1999), 49). Pero sus experimentos –llevados a cabo en la década del 70 empleando técnicas accesibles en ese momento como electroencefalografía y electromiografía asociadas con un osciloscopio de rayos catódicos- mostraron que "Freely voluntary acts are preceded by a specific electrical change in the brain (the 'readiness potential', RP) that begins 550 ms before the act. Human subjects became aware of intention to act 350–400 ms **after** RP starts, but 200 ms. before the motor act. The volitional process is therefore **initiated** unconsciously." (Idem, 47).

unconsciously prepared much longer ahead than previously thought. But we do not know yet where the final decision is made. We need to investigate whether a decision prepared by these brain areas can still be reversed."

El enfoque fisiológico-neural en cuestión puede analizarse como precedencia temporal + correlación constante de variables con un valor de *delay* fijo. En este caso el incremento que se registro en el *delay* temporal –de 200 milisegundos a 10 segundos- refuerza la precedencia temporal, pero dicho refuerzo no puede interpretarse por si mismo como evidencia favorable a una interpretación “fisiologista” de la decisión, a menos que dicha interpretación sea *reforzada* por ME o TI.

Si aceptamos que dicho experimento y las líneas de investigación que surgen en él debieran ser interpretadas en dirección del ME o TI entonces estaríamos diciendo que todo el discurso proléptico + intencional sería completamente falso, del mismo modo que hablamos de la salida y puesta del Sol, o que hablamos de las personalidades y los signos astrológicos; serían falsedades que permiten entender lo que observamos pero que no proveen de conocimiento alguno. Y

en ese contexto el vocabulario PME debería traducirse al vocabulario neurofisiológico de modo que cada término-tipo del vocabulario folk correspondería un teorema de mismidad de significado v.g. “Elección de X sobre Y” debería identificarse estrictamente con “Secuencia de estímulos de la sección X del Núcleo Caudado de acuerdo con los patrones P1... P_n”.

Pero aquí aparecen dos cuestiones metodológicas; por un lado ya señalamos que la correlación *simpliciter* no supone causalidad; pero a ello se agrega que el concepto de racionalidad agencial exige un mínimo de caracterización proléptico-intencional. Este tema lo ha tratado en forma precisa Luisa Montuschi quién resume muy bien el impacto económico de dicha la interpretación “fisiologista” afirmando que estas observaciones experimentales parecen señalar que “... no habría lugar para el libre albedrío...” (Montuschi 2008, p.26-7); pero esto supone un problema porque como la autora señala “La racionalidad práctica presupone el libre albedrío y el proceso de deliberación del agente enfrentado a un conjunto de posibilidades alternativas implícito en el mismo. Si una acción fuera automática y necesariamente generada a partir de creencias, deseos u otros estados mentales

antecedentes, estaríamos totalmente fuera del dominio de la racionalidad, como lo estaría una persona *in the grip of an obsession or an addiction*". (Montuschi; 2006, 8) ¿Pero es esta la única interpretación metateórica que se puede dar a estos resultados?

El carácter problemático de esta cuestión no surge meramente de la pregunta metafísica por el determinismo o el libre-albredismo ontológico –cuestión debatida y en principio aporética- ni del grado de representatividad del experimento. El problema reside en la doble faz teórica y técnico-política de la teoría de la decisión: la elección de un enfoque de primacía de la explicación neurológica justificará un enfoque de primacía de las intervenciones políticas y técnicas mecanicistas, mientras que un enfoque de primacía de la explicación proléptico-intencional justificará un enfoque de intervenciones políticas y técnicas libre-albedristas; en cada caso esto permitirá proveer fondos a ciertas líneas de investigación, formar profesionales que se habitúen a pensar en uno u otro modo, acentuar los programas de acción que

tengan como fin la modificación de estados neurales-intencionales¹⁵.

¹⁵ Una objeción intuitiva a este punto es la siguiente: el conocimiento no debe frenarse por razones ético-políticas, si efectivamente se descubre que la libertad no existe y que la capacidad que tiene un agente de influir voluntariamente en su comportamiento es una ilusión, entonces aunque esto suponga una “herida narcisista” debemos adecuar nuestras instituciones y nuestras prácticas a dicha situación, y no se puede rechazar un conocimiento de hechos apelando exclusivamente a sus efectos políticos. Pero la cuestión es que dicha conclusión no es empírica, sino que surge de un determinado esquema de interpretación y de la compleja dialéctica que se establece entre el material empírico y dichos esquemas, y en ese caso es perfectamente lícito rechazar una conclusión que pretende pasar por empírica, cuando lo que ha sucedido es que se ha *naturalizado* un esquema metateórico. El problema que mencionamos es absolutamente semejante a la que se planteó con la obra Richard J. Herrnstein & Charles Murray *The Bell Curve* (Glencoe, Free Press 1994), en el cual los autores hacían propuestas de política social a partir del supuesto descubrimiento empírico de que el éxito o fracaso social estaba determinado por el CI y que este era biológico, hereditario e inmodificable (v.g. las políticas proactivas hacia los negros eran un gasto de dinero, porque los negros eran menos inteligentes que los blancos, dado que sus test de CI daban sistemáticamente más bajos). Dada la magnitud de las afirmaciones fácticas y de las propuestas políticas que aparecían en ese texto, la American Psychological Association ofreció un dictamen institucional “Intelligence: Knowns and Unknowns” (*American Psychologist* 51 (1996) 77-101) donde se decía que “Psychometricians have successfully measured a wide range of abilities, distinct from one another and yet intercorrelated. The complex relations among those abilities can be described in many ways. Some theorists focus on the variance that all such abilities have in common, which Spearman termed g (“general intelligence”); others prefer to describe the same manifold with a set of partially independent factors; still others opt for a multifactorial description with factors hierarchically arranged and something like g at the top. Standardized intelligence test scores (“IQs”), which reflect a person’s standing in relation to his or her age cohort, are based on tests that tap a number of different abilities. Recent studies have found that these scores are also correlated

Pero la interpretación desde un léxico/ontología FH permite una integración no conflictiva de los resultados experimentales y los niveles jerárquicos con primacía del componente proléptico-intencional. El mismo Benjamin Libet rechazó una interpretación estrictamente reduccionista y defendió un rol para los niveles proléptico-intencionales (vid.

with information processing speed in certain experimental paradigms (choice reaction time, inspection time, evoked brain potentials, etc.), *but the meaning of those correlations is far from clear.* (95-6) El punto es que las mismas correlaciones reciben interpretaciones completamente diferentes y hasta contradictorias, y no se trata de ciertos mecanismos teóricos postulados sino que la declaración es clara sobre la cuestión: se trata de qué significado tengan dichas correlaciones, y no de qué hipótesis teórica permita explicarlas.

Hay otros casos análogos a este, en que se presentan diferentes esquemas metateóricos que llevan a diferentes políticas e intervenciones v.g.:

- (I) Los psicoanalistas se han negado y se niegan a interactuar con los enfoques neurales de la enfermedad mental, y en particular han sido son completamente hostiles a cualquier idea de neurofarmacología y dedican sus esfuerzos teóricos y sus recomendaciones de política de salud mental a enfoques mínimamente psiquiátricos.
- (II) Los ambientalistas radicales en criminología sostienen que las explicaciones individualistas y psicológicas del comportamiento criminal son irrelevantes y dedican sus esfuerzos teóricos y sus recomendaciones de política criminal a enfoques abolicionistas del derecho penal.

En ambos casos no se trata de una mera oposición factual sobre verdades o falsedades sino fundamentalmente de una oposición conceptual —metafísica- y práctica sobre que aspectos de la realidad serán relevantes y cuales serán irrelevantes.

Libet, 2001 & 2003). Este autor introduce una hipótesis claramente proléptica como tesis necesaria en una teoría de la decisión que integre resultados neurológicos: la tesis del veto consciente, tesis que es incompatible con las lecturas desde TI o ME. Libet sostiene que la interpretación usual de sus propios resultados es que “Unconscious initiation of the voluntary process appeared to mean that conscious free will could not actually ‘tell’ the brain to begin its preparation to carry out a voluntary act.” (Libet, 2003, 24). Si la volición consciente no es un efector de la acción entonces la volición sería etiológicamente irrelevante e irrelevante sin más. Pero Libet señala que los mismos experimentos mostraron que “...the conscious wish to act (W) preceded the final motor act by about 200 msec (or 150 msec when corrected for a –50 msec bias in the reporting process). That provided an opportunity for the conscious will to control the outcome of the volitional process. It could do that by providing a possibly necessary trigger for the process to proceed to completion in a motor act; there is, however, no evidence for such a mechanism. Or, the conscious will could block or ‘veto’ the process, resulting in no motor act.” (idem).

En este párrafo Libet sostiene dos cosas diferentes:

- (i) Un resultado experimental: existen una secuencia de señales compleja, que incluye un momento W posterior al primer efector no conciente-volitivo y es parte del conjunto de efectores que producen el acto motor.
- (ii) Una interpretación teórica, que señala que dicho momento W tiene una función de veto o censura volitiva sobre la secuencia Efector primario → W → Movimiento activo.

En ellos opera un ***metateoretical framework*** no reductivo, ya que Libet introduce una dimensión proléptico-intencional en la elección de una interpretación de dichos resultados experimentales. Y la interpretación funcional-homuncular es perfectamente intuitiva: los niveles neurales homuncularmente con los niveles agenciales, y la interacción entre proceso neural y efector volitivo se analiza en términos de los tipos institucionales que postula FH.

(B) La irrelevancia del agente/decididor. El sesgo de las interpretaciones informadas por TI o ME parece hacer innecesario el concepto de agente/decididor; al menos en esa dirección parecen aportar los resultados que expone Daniel M. Wegner. Este autor ha señalado que hay una desconexión etiológica entre nuestros reportes de decisión – aquellos que informan de la sensación conciente que nuestros deseos y creencias son efectores de nuestras decisiones o son evidencia favorable para suponer que provocamos la decisión- y el comportamiento decisional efectivo. En el *precis* de su obra principal el autor sostiene que: “This book brings these cases together with research evidence from laboratories in psychology and neuroscience to explore a theory of *apparent mental causation*. According to this theory, when a thought appears in consciousness just prior to an action, is consistent with the action, and appears exclusive of salient alternative causes of the action, we experience conscious will and ascribe authorship to ourselves for the action. Experiences of conscious will thus arise from processes whereby the mind interprets itself – not from processes whereby mind creates action. Conscious will, in this view, is an indication that we *think* we have caused an

action, not a revelation of the causal sequence by which the action was produced.” (Wegner, 2004, 1).

Evaluamos a un agente elector por los informes que nos da, por la consistencia que ellos tienen con su comportamiento; por la presencia de normas y estimaciones; en cada caso los reportes de creencias, valoraciones, deseos y demás son relevantes para las evaluaciones y predicciones del agente elector. Pero si los procesos agenciales son meramente “...indication that we *think* we have caused an action, not a revelation of the causal sequence by which the action was produced” entonces los reportes subjetivos de acción y decisión no serán evidencia de nada que hagamos o podamos hacer sino que serán más parecidos al ruido sin que tenga ningún interés su contenido pues sea cual sea, no forma parte de la trama etiológica del comportamiento del agente elector; dicho de manera cruda, de acuerdo con este análisis cuando evaluamos decisiones evaluamos cerebros y la agencia personal –sería considerada un mero epifenómeno de los dispositivos causales sub-personales de dichos cerebros. Y en ese caso cualquier proyecto de teoría de la decisión con un componente proléptico-intencional sería ciencia pretérita en modo analogable a la situación de la

química pre-lavoasariana respecto de la química atómico-molecular.

Pero la conclusión de este autor puede ser reinterpretada en un esquema FH; recordemos que este autor sostuvo que "...indication that we *think* we have caused an action, not a revelation of the causal sequence by which the action was produced". De acuerdo con este esquema es claro que nuestros reportes no darían cuenta de los efectores de ciertos niveles jerárquicos inferiores, pero estos niveles estarían homuncularmente vinculados con la conducta molar, por ello aunque los informes de voluntad no sean efectores de los procesos neurales, ellos contienen homuncularmente dicho carácter efector, junto a otros rasgos. Dicho de otra manera, sí evaluamos cerebros, pero cerebros en cuerpos con conducta molar adaptable a entornos y con roles sociales, y los informes de voluntad involucran estos y son efectores no del comportamiento del cerebro sino de la conducta molar en entornos normativos.

(C) La vacuidad explicativa del programa reductivista.

No solo se pueden reinterpretar los resultados experimentales en un modo no reduccionista –con el auxilio de FH- sino que es posible señalar que **todo** programa reduccionista –sea ME o TI o alguna variedad o alternativa a ellos- cae en un vacío explicativo, y que aún si se sostuviera una perspectiva materialista, el léxico/ontología proléptico-intencional es indispensable.

Supongamos una teoría biológica sobre algún fragmento del comportamiento humano vg. una teoría genética que explique la adicción a las drogas y que logre solucionar el problema de la evidencia falsamente favorable: “With these caveats in mind <i.e. el despeje de casos de alcoholismo social o socialmente sesgado que parecieran deberse a influencias biológicas>, in the genetic hypothesis an individual would enter life with a certain level of a genetically influenced biological predisposition toward alcoholism or drug abuse. It is probable that multiple genes are involved or that other factors affect the strength of the actions of a particular gene (i.e., incomplete penetrance of the gene). If the disorder is polygenic (i.e., involving more than one gene) there is probably a combination of genes which might predispose an individual to alcoholism (e.g., the possibility of getting a

different level of intoxication when drinking or an unusual effect of alcohol on anxiety, etc.) and some which might help to protect a person from demonstrating alcoholism (e.g., becoming very ill at even low alcohol doses). The person, then, could go through a variety of life events and stresses, some of which would predispose him or her to alcoholism (e.g., working in a heavy-drinking environment, such as the armed services) and others which would protect the person from demonstrating the predisposition (e.g., being a woman in a society with heavy proscriptions against drinking for women). The final alcoholic picture would depend upon the balance between the positive and negative genetic effects interacting with the positive and negative environmental factors." (Schuckit, Marc A., 1980. 299).

Estos eventos ambientales pueden ser instrucciones, creencias morales o religiosas, normas de etiqueta, reglas de urbanismo, consejos prudenciales sobre una vida sana o alguna combinación de ellos; en muchos casos estos elementos proléptico-intencionales operan como restrictores – así lo reconoce explícitamente Marc Schuckit- en exacta correspondencia con los eventos W que señalaba B. Libett.

Así la forma general de las teorías genéticas del comportamiento será la siguiente:

Efactor primario → **W₁** Efactor _{nn} → **W_{nn}** → **Ejecución**

el efector primario será "Configuración genética G", cuya presentación puede asumirse estrictamente orgánica-neural, pero ello no será el caso de los efectores W₁...W_n que incluirán un elemento proléptico-intencional –y que desde la perspectiva de FH, tendrán un compromiso homuncular entre sí.

Dado que el comportamiento molar no es un dato sino un constructo multidimensional, y siendo este carácter sistémico lo que hace tan difícil aislar los aspectos genéticos las teorías genéticas no suponen una alternativa al léxico/ontología proléptico-intencional sino una ampliación de los esquemas de explicación y de tecnologías-políticas que exige dicho componente pues solo en término proléptico-intencionales se puede caracterizar el complejo sistémico de efectores aún en el caso que el efector primario sea "Configuración genética G".

Peor aún, aún en el caso de comportamientos en los que se reconoce la presencia de efectores genéticos, su curación o contención no solo se hace a nivel biológico –por medio de psicofármacos- sino a nivel dinámico, conductual y vincular...Aún si el alcoholismo fuera estrictamente un tipo de comportamiento genéticamente determinado, su tratamiento – por medio de efectores secundarios como W- hace indispensable la intervención de una dimensión proléptico-intencional.

VII. A modo de conclusión abierta.

El examen anterior, intentó presentar el Funcionalismo Homuncular como un *methateoretical framework* relevante para la investigación científico-filosófica de la interfaz entre las ciencias económicas y las neurociencias, si bien este es un dominio en el que siquiera hay algún terreno común. El modelo de *homo oeconomicus* supone un dispositivo motivador nulo, pues desde el punto de vista del consenso robbinsoniano el agente económico se construye con una caja negra con entradas de preferencias y salidas de elecciones de satisfactores escasos (por esa razón la economía robbinsoniana se puede aplicar a comunidades de

mamíferos o de autómatas); en ese sentido la neurociencia es irrelevante sustantivamente y solo sirve para el ajuste de las predicciones, o sea solo le sirve al economista como un insumo teórico más. Sin embargo este enfoque estrictamente dualista –hay economía y hay neurociencia y solo tienen relaciones instrumentales- puede ser contestado desde una visión integradora, y en ese sentido el Funcionalismo Homuncular brinda una excelente herramienta para dicha integración sustantiva y no solo instrumental.

Y la cosa es particularmente relevante cuando los experimentos y las paradojas de la teoría de la decisión llevan a abrir la caja negra del decidor racional por un agente encarnado no tan racional o mínimamente racional o sesgadamente racional. En ese sentido las neurociencias parecen aportar ciertas hipótesis centrales en nuevas teorías de la decisión, y nuevamente el Funcionalismo Homuncular parece relevante para la integración de dichas hipótesis y las hipótesis estándares. En ese sentido considero que la asunción de un punto de vista análogo o informado en el Funcionalismo Homuncular es una tarea que la comunidad de economistas experimentales y de psicólogos de la decisión deben encarar.

Bibliografía

- Bentham Jeremy (1781/2000): *An Introduction to the Principles of Morals and Legislation*, Kitchener, Batoche Books.
- Block; Ned (1980/1993): "Troubles with functionalism", en *Readings in philosophy and cognitive science*, ps. 231-253 Cambridge, MA The Mit Press
- Brentano, Franz (1935/ 1874) *Psicología desde un punto de vista empírico* Traducción de José Gaos, Madrid *Revista de Occidente*.
- Commons, John, (1931): "Institutional Economics" *American Economic Review*, 21 (648-657)
- Churchland, Paul (1981) "Eliminative Materialism and the Propositional Attitudes" *The Journal of Philosophy* 78, 67-90
- Davidson, Donald (1997/1990) "Estructura y contenido de la verdad"; en Juan Antonio Nicolás y María José Frápoli (eds.), *Teorías de la verdad en el siglo XX*. Tecnos, Madrid.

- Hardy-Vallée, Benoit (2007) “Decision-Making: A Neuroeconomic Perspective” *Philosophy Compass* 2 (939-953)
- Haynes, John-Dylan *et. al.* (2008) “Unconscious determinants of free decisions in the human brain” *Nature Neuroscience* 11, 543 – 545.
- Haynes, J.-D. & Cleeremans, A. (1999). “Correlating consciousness: A view from empirical science”. *Revue Internationale de Philosophie*, 3, 3 87-420; también en <http://srsc-mac1.ulb.ac.be/axcWWW/papers/pdf/98-NCC.pdf>.
- Lettieri, Dan J. & al. (eds.) (1980), *Theories on Drug Abuse, Selected Contemporary Perspectives* NIDA Research Monograph 30, National Institute on Drug Abuse, Maryland.
- Lewis, David & Bridger, Darren (2005)“Market Researchers make Increasing use of Brain Imaging”, *Advances in Clinical Neuroscience and Rehabilitation* 5/3 36-37
- Libet, Benjamin (2003) “Can Conscious Experience Affect Brain Activity?” *Journal of Consciousness Studies*, 10, 24–28.

- Libet Benjamin & Haggard Patrick (2001) "Conscious Intention and Brain Activity" *Journal of Consciousness Studies*, **8** 47–63.
- Libet, Benjamin (1999) "Do We Have Free Will?" *Journal of Consciousness Studies*, **6**, 47–57.
- Lycan, William G. (1995): *Consciousness*, Cambridge, MA; The MIT Press Pbk ed
- McClure SM et al.(2004) "Neural Correlates of Behavioral Preference for Culturally Familiar Drinks", *Neuron*;44:379–87
- Mill; John Stuart (1844) *Essays on some unsettled questions of political economy* John W. Parker, London.
- Montuschi; Luisa, (2006) *El libre albedrío y los conceptos de racionalidad y de elección racional en la ciencia económica* Universidad del Cema, Documentos de Trabajo, N° 324, Buenos Aires
- -----, (2008) *La libertad, el libre albedrío y el enfoque de las capacidades: el desarrollo humano como libertad*, Universidad del Cema, Documentos de Trabajo, N° 372, Buenos Aires.

- Place, U.T (1956) "Is consciousness a brain process?" *British Journal of Psychology* 47, 44-50.
- Popper, Karl (1995), *The Myth of the Framework. In Defense of Science and Rationality*, Oxford, Routledge
- Putnam, Hilary (1975/1966) "Mind and Machines" en *Mind, Language and Reality*, C.U.P, Cambridge, 362-385.
- Schuckit, Marc A. (1980) "A Theory of Alcohol and Drug Abuse: A Genetic Approach" en Lettieri et. al., 297-308
- Wegner, Daniel M. (2004) "Précis of *The illusion of conscious will*", *Behavioral and Brain Sciences* (2004) 27, 1-46.

¿DE QUÉ SIRVE MAXIMIZAR LA UTILIDAD ESPERADA?

Gustavo Marqués

I. Teoría estándar de la decisión

I.1. Introducción

En condiciones de *certeza* los individuos conocen con seguridad el resultado de sus elecciones y tienen preferencias definidas sobre ellas. Por ejemplo, si saben que la acción A_1 garantiza que obtendrán el objeto X_1 y la acción A_2 el objeto X_2 , y prefieren X_2 a X_1 , es racional que emprendan la acción A_2 . Llamaremos *canastas* (de bienes) a los objetos de elección obtenibles en estas condiciones. La noción estándar de *elección racional* relevante para la teoría económica está ligada centralmente a la de *preferencia*. En una primera aproximación podemos decir que los individuos eligen aquello que prefieren y es racional que procedan de esta manera. Pero la teoría es algo más exigente: impone

restricciones al conjunto de preferencias, las cuales consisten en la satisfacción de los axiomas de completitud, transitividad y continuidad. Cuando el conjunto de preferencias está estructurado de esta manera, se dice que están *bien formadas* y pueden ser ordenadas conformando un *orden débil*, es decir, un ranking completo y transitivo. Las preferencias que poseen esta estructura pueden ser representadas por una *función de utilidad ordinal*. La elección racional consiste en *maximizar esa función*, es decir, no elegir una canasta determinada X estando disponible otra canasta Y mejor rankeada. Cuando un individuo elige siguiendo esta pauta se comporta racionalmente (maximiza).

Bajo condiciones de *riesgo* (e incertidumbre) los individuos afrontan una situación de elección diferente a la recién mencionada. Ahora deben elegir entre pares de opciones del siguiente tipo:

Opción 1: obtener una heladera y un microondas con probabilidad 0.2 o, en caso contrario, nada;

Opción 2: obtener una heladera con probabilidad 0.25, o, en caso contrario, nada.

Opciones del tipo mencionado son denominadas *loterías*. Una lotería es un conjunto conformado por uno o más resultados y las probabilidades asociadas a cada uno de ellosⁱ. Si el sujeto elige la opción 2 puede obtener una heladera, pero corre el riesgo de no obtener nada. Las variables relevantes en la elección riesgosa son los *resultados* que podrían obtenerse de elegir una cierta lotería (por ejemplo, una heladera o nada, en la opción 2) y las *creencias* acerca de la *probabilidad* de obtener dichos resultados (0.25 y 0.75, respectivamente). Como generalmente se asume que dichas creencias son *correctas*, puede decirse simplemente que bajo riesgo los conceptos clave son los de *resultado* y *probabilidad* asociada a cada resultado.

La teoría estándar de la decisión incorpora la idea Bernoulliana de que, una vez que han sido asignados valores adecuados, las probabilidades y las utilidades asociados a los resultados de una lotería pueden ser multiplicados entre sí y estos productos adicionadosⁱⁱ. Esta sumatoria de productos es lo que Friedman y Savage designan como Hipótesis de la Utilidad Esperadaⁱⁱⁱ. La expresamos así:

$$\sum_{i=1}^n p_i \cdot u(x_i)$$

[EUH]

p_i es la probabilidad del premio i y $u(x_i)$ es su utilidad.

Este procedimiento matemático brindaría una representación adecuada de la *utilidad* que los individuos pueden esperar en caso de elegir una determinada lotería. El concepto de “utilidad” tal como es usado aquí está despojado completamente de toda connotación psicológica y representa un mero índice que permite conformar un ranking con las diferentes opciones. Volviendo a las opciones mencionadas anteriormente, si fuera posible asignar valores numéricos a las utilidades que proporcionan los premios de una lotería^{iv} y, por ejemplo, la utilidad de una heladera adquiere para una persona el valor 0.5 y el de una heladera y un microondas el valor 0.7, sus utilidades esperadas de las opciones mencionadas arriba serán las siguientes:

Opción 1: $(0.7 \times 0.2) = 0.14$

Opción 2: $(0.5 \times 0.25) = 0.125$

La acción racional consiste en maximizar la utilidad *esperada*. Maximizar es simplemente elegir aquella lotería mejor rankeada (a la que ha sido asignado el número mayor). Lo que se maximiza es la utilidad *esperada*, no la utilidad segura, porque recuérdese que en cualquiera de los dos casos es posible que la persona no obtenga nada. El sujeto racional elegirá la primer opción, cuya utilidad esperada es mayor. Sobre estas mismas bases se desarrollará posteriormente casi toda versión de la teoría de la decisión relevante para la economía.

1. 2. Teoría de la Utilidad Esperada (EUT)

En 1944 John von Neuman y Oskar Morgenstern proponen una teoría axiomática de la elección bajo *riesgo*, que implica la hipótesis Bernoulliana de la utilidad esperada.

Los axiomas son los ya mencionados para el caso de certidumbre (completitud, transitividad y continuidad), más un axioma adicional: el axioma de independencia. EUT contiene también supuestos vitales, como el de reducción de loterías complejas a simples y los supuestos de invariancia. Si las

preferencias sobre loterías satisfacen dichos axiomas y supuestos, se dice que están bien formadas y son representables mediante una función de utilidad que posee la propiedad de la utilidad esperada y es única hasta una transformación afín positiva. La elección racional bajo riesgo e incertidumbre consiste en maximizar dicha función.

Hay varias razones por las que EUT ejerció un fuerte atractivo sobre los economistas. En primer lugar, es una teoría mínima de la racionalidad: todo lo que exige es

- (a) que las (creencias acerca de) probabilidades y las preferencias sobre loterías satisfagan las condiciones de *coherencia* que imponen los axiomas.
- (b) que no se seleccione una lotería habiendo disponible otra mejor rankeada.

Notar que la racionalidad incorporada a EUT es, pues, meramente coherentista, no sustantiva. Además, en condiciones de riesgo, se releva a los individuos de ser consistentes respecto de las probabilidades, debido a que éstas ya están dadas (es un dato para ellos). Sólo se requiere que sean *consistentes* respecto de la *asignación de utilidades*

implicaciones para la segunda: ahora, para ser consistentes, *deberían* preferir a D sobre C. Deberían hacerlo porque C y D se obtienen de multiplicar A y B por 0.25, lo que significa que la segunda elección es en definitiva idéntica a la primera. Esto transforma a EUT en una teoría *empírica*, aplicable al dominio de los seres humanos. En la situación de elección descrita, si se sabe que los individuos han elegido a B sobre A, la teoría *predice* que elegirán a D sobre C.

Supongamos que los individuos se comportan de manera *opuesta* a la prescrita en la teoría y que habiendo elegido B en la opción inicial eligen C en la subsiguiente. Que prefieren B a A significa que

$$u(x_2) > 0.8 u(x_1) \quad (I)$$

Que prefieren C a D implica que

$$0.2 u(x_1) > 0.25 u(x_2) \quad (II)$$

Multiplicando ambos miembros de (I) por 0.25 tenemos

$$0.25 u(x_2) > 0.2 u(x_1)$$

lo que contradice lo expresado en (II).

El patrón de elección que acabamos de describir expresa una *inconsistencia* en la asignación de utilidades a los resultados de las loterías (es decir, las elecciones efectuadas por los agentes reflejan una asignación de utilidades inconsistente). Comprometerse con que la teoría es aplicable (exitosamente) al dominio de los humanos es apostar a que los humanos poseen un conjunto de asignaciones de utilidad consistente. Y es por esta razón que sus elecciones pueden ser predichas. La consistencia de sus elecciones es condición necesaria de su capacidad predictiva. Contar con capacidad predictiva (o, lo que es lo mismo, contenido empírico), era considerada una propiedad *necesaria* de las teorías científicas, aunque más no fuere por su vinculación con su contrastación empírica.

Cabe remarcar, sin embargo, que inicialmente, EUT fue considerada una teoría *normativa* de la decisión, aunque sus prescripciones eran un tanto limitadas. Es importante especificar el sentido restringido en que EUT instruye a los tomadores de decisiones. Una vez que las diferentes opciones han sido volcadas en el formato de EUH, la

concepción de “normativo” involucrada en la visión estándar se reduce a lo siguiente: a) los agentes deben ser consistentes en sus asignaciones de utilidad a los resultados componentes de las loterías; b) los agentes deben maximizar su utilidad esperada. (a) y (b) son *relativamente* independientes, ya que un sujeto podría cumplimentar (a), pero no (b). El fracaso en maximizar su utilidad no implica su fracaso en asignar utilidades consistentemente. Pero, a la inversa, si fracasa en (a) no puede tener éxito en (b): la consistencia en la asignación de utilidades a los resultados es *condición necesaria* de la maximización.

Es importante comprender qué implica exactamente la condición (a). Retomando la elección en dos etapas descrita anteriormente, EUT no indica al sujeto cómo debe elegir entre A o B, ni cómo hacerlo entre C y D. Pero si el sujeto ha elegido B (A) en la primer elección, debe elegir D (C) en la segunda. Todo el contenido normativo de la teoría consiste en señalarle al sujeto su inconsistencia en el conjunto de elecciones. La inconsistencia debe ser reparada, pero la teoría no indica cómo hacerlo. El individuo que ha elegido B en la primera elección y C en la segunda puede rectificarse de tres maneras posibles: 1) manteniendo su preferencia por

B, pero eligiendo ahora a D; 2) manteniendo su preferencia por C, pero eligiendo ahora a A; 3) rectificando ambas decisiones iniciales y declarándose ahora indiferente en ambas elecciones. Por ende, todo lo que la teoría prescribe, en primer lugar, es que se repare la inconsistencia. Una vez que el sujeto ha asignado a los resultados una asignación de utilidades consistente, la segunda prescripción de EUT es que se maximice la utilidad esperada (es decir, que se elija aquella lotería cuya utilidad esperada es mayor).

El sueño dorado de contar con una teoría de la decisión en condiciones de riesgo que fuera adecuada, a la vez, desde el punto de vista descriptivo y normativo, parece haber naufragado. Numerosos experimentos realizados durante la segunda mitad del siglo XX mostraron que EUT era insatisfactoria desde el punto de vista descriptivo: los agentes no se comportaban *como si* maximizaran su utilidad esperada^v. La reacción ante esta circunstancia consistió en privilegiar el objetivo de “racionalizar” los fenómenos, dando entera libertad para modificar los axiomas de la teoría de manera de obtener los resultados apetecidos sin importar lo arbitrario de los supuestos invocados para ello. De esta manera, se introdujeron cambios, primero en su periferia

(Axioma de Independencia)^{vi} y luego en su núcleo mismo (transitividad)^{vii}. Ingentes cantidades de talento e ingenio fueron puestos al servicio de “salvar” los fenómenos^{viii}. El objetivo de todas estas modificaciones es introducir transformaciones sobre las dos variables clave de la ecuación

básica $\sum_{i=1}^n p_i \cdot u(x_i)$ –probabilidades y utilidades– con el propósito de acomodar las anomalías.

Pese a que de esta manera han sido alcanzados ciertos logros importantes, la estrategia “descriptivista” no parece ser completamente satisfactoria en referencia a sus propios objetivos. Es un hecho establecido que en numerosas circunstancias los agentes se comportan *inconsistentemente* respecto de la teoría estándar, lo que significa que *no* maximizan su utilidad esperada. Ciertos comportamientos bien establecidos bajo condiciones experimentales no han sido explicados aún por ninguna versión de la teoría económica de la decisión^{ix}. Se verifica un exceso de información empírica en relación a las capacidades sistematizadoras de la teoría disponible. Y no hay señales de que esta brecha tienda a reducirse.

Hay un tipo particular de inconsistencia que tiene graves implicaciones para la teoría estándar de la decisión. Algunas decisiones son sensibles a lo que Kahneman y Tversky (1981, 1986) denominan “framing effects”, que consiste en que “formulaciones equivalentes del mismo problema de decisión arrojan sistemáticamente preferencias diferentes”. Por ejemplo, la elección entre dos terapias alternativas recibe diferentes respuestas si exactamente el mismo problema de elección es descrito en términos de vidas que se salvan o vidas que se pierden. Como los autores notan, es evidente que en estos casos “la gente elige entre descripciones de opciones más que entre las opciones mismas” (Tversky, 1999, p. 180). Si hubiera una relación uno a uno entre cada situación de elección objetiva y su correspondiente descripción, no existiría ningún problema: sería cuestión de convención o comodidad expresiva decir que los individuos eligen entre un conjunto de opciones dado o la descripción de las mismas. Los “framing” muestran que este no es el caso. Pero aún así, admitiendo la posibilidad de ofrecer diferentes descripciones alternativas de un mismo conjunto de opciones, tampoco se presentaría dificultad alguna si aquellas pudieran ser reducidas a una única forma canónica, equivalente a todas ellas. El problema reside en que no existen

mecanismos psicológicos que realicen de manera espontánea la tarea de reducir descripciones diferentes a una única forma representativa. Aunque los sujetos interpretan sin dificultad que enunciados expresados en voz activa y pasiva son equivalentes en contenido (es decir, que “Juan pintó esta mesa” y “Esta mesa fue pintada por Juan” transmiten la misma información), no advierten que las diversas descripciones de un mismo problema son equivalentes^x.

Existen diversos tipos de framing effects, que han mostrado ser un resultado robusto y extendido (Kahneman y Tversky, 1981, 1986). Su relevancia, en relación a lo que se discute en este artículo, consiste en que violan el supuesto de invariancia, que es *condición necesaria* para que la noción de utilidad resulte medible. Su violación significa que los individuos *no* poseen un orden bien definido de preferencias que pueda ser representado mediante una función de utilidad^{xi}. Se pierde, en consecuencia, lo que fue considerado un logro extraordinariamente importante, que distinguía a la teoría económica del resto de las ciencias sociales.

Los framing effects también dan por tierra con las aspiraciones *normativas* de EUT en cuanto a consistencia se refiere. Dado un marco de elección (es decir, una cierta descripción de la situación de elección que enfrentan) los individuos podrían comportarse inconsistentemente al comienzo, pero, al ser advertidos de ello, introducir rectificaciones de manera de restaurar la consistencia en sus elecciones. Pero en la medida en que la misma situación de elección puede ser descrita de diferentes maneras, si éstas no son reducibles a un marco común (una forma canónica, en términos de Kahneman), se carece de mecanismos capaces de asegurar la consistencia de las elecciones ante las diversas descripciones. No contarían pues con un único conjunto de preferencias bien formadas, que se revelarían en sus elecciones. La instrucción: “sea consistente en su asignación de utilidades a los resultados” es impracticable para los individuos (excepto al interior de un marco de elección). Esto es menos que lo aspirado (o presupuesto) por EUT.

II. Nuevo proyecto. La soberanía del decisor y el supuesto de correspondencia entre preferencias y utilidades

Pero supongamos, al sólo efecto de proseguir la discusión, que esta situación pudiera solucionarse, ya sea porque los framing effects no tienen sobre las decisiones el impacto que se les atribuye, o los sujetos, de alguna manera, aprendieran a reducir las diversas descripciones de una situación de elección a una única forma representativa. Esto significaría que al seleccionar entre diferentes loterías los individuos *pueden* ser consistentes en sus asignaciones de utilidad y contar con una función de utilidad definida sobre loterías. El papel de EUT, al identificar la inconsistencia e instruir en que ésta sea reparada, resultaría importante en estas condiciones. Y restaurada la posibilidad de comportamiento consistente, la segunda instrucción de EUT, que exige maximizar la utilidad esperada, entra en vigencia. Podemos ahora preguntarnos ¿por qué razones EUT considera que la conducta maximizadora es racional?

Una buena teoría económica de la decisión debe satisfacer las necesidades de dos clases de actores diferentes: las de

los teóricos y las de los agentes. Desde el punto de vista del *teórico*, se requiere que EUT sea *descriptivamente adecuada* para racionalizar o predecir patrones de conducta en situaciones riesgosas o inciertas. Aparentemente, una mayoría de especialistas –entre los que encuentran autores tan diversos como Friedman, Kahneman y Tversky y Starmer– se inclina por esta visión de la teoría de la decisión. Sin embargo, de ser factible, el éxito del proyecto descriptivista podría ser irrelevante desde la perspectiva de los agentes. EUT prescribe que éstos *maximicen su utilidad esperada*. ¿Por qué sería esto importante *para ellos*? De que les serviría a los *individuos* maximizar su utilidad *esperada*?

Una teoría *normativa* de la decisión recomienda a los agentes determinadas pautas de conducta, no con el propósito de manipularlos, sino *porque* asume que sus instrucciones son *valiosas para quienes las ponen en práctica*. He aquí una cierta ambigüedad que refleja dos concepciones diferentes, potencialmente conflictivas, del servicio que la teoría debe prestar a los individuos. De una parte, existe una larga tradición en economía que piensa a la ciencia económica como un saber superior al de los propios actores. En este sentido el papel normativo de la teoría consiste en contribuir a

mejorar las elecciones de los agentes y, eventualmente, sugerir *rectificaciones*. En buena medida, este papel es el que analizamos anteriormente: consiste en detectar inconsistencias en la elección y demandar rectificaciones. De otra parte, en cuanto ligada estrechamente a la doctrina individualista, EUT instruye qué *deben* hacer los agentes para alcanzar los *propósitos que ellos mismos desean obtener*. Esto convierte a la teoría estándar de la decisión es un instrumento puesto al servicio de lo que podríamos llamar la *soberanía del decisor*: contribuye a que éste pueda alcanzar *sus* objetivos. En consecuencia, una teoría normativa adecuada, tiene validez *subjetiva* y *objetiva*. Desde el punto de vista subjetivo debe contribuir a que el decisor pueda ejercer satisfactoriamente su soberanía. Desde el punto de vista objetivo debe ser capaz de rectificar sus decisiones, si fuera necesario. Ambas propiedades son necesarias: la primera para que las elecciones que recomienda sean *relevantes* desde la perspectiva de los propios actores, y la segunda para que sean (objetivamente) *racionales*. En cuanto teoría normativa, la teoría estándar ha hecho hincapié en su papel *objetivo* (del cual nos hemos ocupado brevemente en la sección anterior), pero ha descuidado por completo el aspecto *subjetivo* (que consideraremos a continuación).

La teoría de la decisión económica estándar ha procurado eliminar por completo toda referencia a estados mentales o psicológicos. En las formulaciones usuales, la *utilidad* se transformó en un mero índice numérico, sin conexión con la satisfacción que el consumo de un bien produce en los individuos^{xii}. Para restituir de manera plena la perspectiva de los propios agentes se necesita dejar a un lado la concepción estándar de “utilidad” y reintroducir la noción “clásica” de utilidad, entendida en el sentido hedonista (psicológico) del término. La noción circunscripta de *utilidad*, como mero índice, respondió al propósito de *medir* la utilidad y al hecho de que la utilidad, en el sentido clásico del término (sensación de placer o displacer), no parecía medible según las pautas de la teoría de la medición establecida. Este propósito fue dictado por la visión de la ciencia vigente en los años 50. Pero, si como sugiere nuestro análisis previo, el proyecto estándar ha fracasado: los individuos no poseen un orden de preferencias bien determinado sobre loterías, susceptible de ser “elicitado” e indexado mediante una función de utilidad (Tversky, 1999; Kahneman y Sugden, 2005), ¿por qué continuar absteniéndose de utilizar el término “utilidad” en el sentido clásico, cuyo referente es, después de todo, tan *inmedible* como el de la noción estándar?^{xiii} ¿Qué propósito

tiene seguir negando la conexión entre “utilidad” y “satisfacción” en el sentido psicológico? La posición que adoptamos en este trabajo se ve reforzada por dos razones, vinculadas entre sí.

En primer lugar, es interesante notar que el esfuerzo por dotar de carácter científico a la economía generó una ruptura entre el significado técnico de “utilidad” y el significado ordinario que usualmente se asigna a este término (Es en este sentido que Friedman, 1955, lamenta la continuidad terminológica). Sin embargo, a pesar de los intentos por desarrollar un discurso formal que permita separar a la economía de la psicología, una vez que han traspasado las puertas del aula o el seminario los significados temporariamente bloqueados vuelven una y otra vez en el discurso informal de los economistas y teóricos de la decisión.

A su vez, la vigencia (informal) de la concepción psicológica de utilidad en el discurso económico es entendible si se presta atención al siguiente argumento *abductivo*. ¿Qué es lo que *explica* que los sujetos se interesen en “maximizar” la utilidad de las opciones que enfrentan? ¿Por qué prestan

atención a las “utilidades”, después de todo? La respuesta (o, al menos, una de ellas) es que prevalece en sus mentes el contenido (psicológico), nunca definitivamente exorcizado, que liga “utilidad” y “satisfacción” (en el sentido de experiencia hedonística). Si no restituyeran el significado latente a la noción de “utilidad”, la máxima “Maximice su utilidad” carecería de toda relevancia para ellos. Ocurre en este caso lo mismo que con muchos desarrollos teóricos, que requieren de un parafraseo informal en términos de “satisfacción” para tornarse comprensibles y relevantes.

En atención a lo expuesto, en este trabajo no adoptamos la decisión quijotesca, propia de la ortodoxia, que priva a la teoría de la decisión individual de conexión con los objetivos y deseos de los individuos, y basamos nuestro análisis en la noción de utilidad *clásica*, entendida como experiencia hedonística consistente en la satisfacción de un deseo. Este enfoque permite reivindicar la jerga informal de los economistas, que expulsan de su discurso los contenidos psicológicos cuando desean ser *precisos*, pero los reintroducen de manera subrepticia y coloquial cuando procuran exhibir la *relevancia* del rigor de sus demostraciones.

Una vez que el interés por los objetivos del propio agente y la concepción clásica de utilidad, con su dimensión hedonista de satisfacción, son restituidos a la teoría, puede advertirse que la teoría estándar es *normativamente inadecuada* en el sentido *subjetivo* elucidado anteriormente. Es decir, es ineficaz para contribuir a que el decisor pueda ejercer exitosamente su soberanía. Disimula su ineficacia porque incorpora, de manera tácita y a – crítica, una suposición (sustantiva) fundamental: la *correspondencia entre preferencias y utilidades* (entendidas, estas últimas, en el sentido clásico del término).

Las diferentes versiones de la teoría de la decisión estándar han modificado uno u otro de los axiomas originales, incluyendo a algunos de los más centrales, pero todas han retenido la presunción de que las preferencias que los individuos tienen sobre los diversos resultados (premios) de las loterías *se corresponden* perfectamente con las utilidades (satisfacciones) que éstos les reportarán una vez que los objetos sean consumidos. Denominaremos a esta tesis Supuesto de Correspondencia (SC) y lo formularemos de esta manera:

SC: las preferencias sobre bienes de los agentes son un buen indicador (en particular, para ellos mismos) de la utilidad (satisfacción) que dichos bienes les prestarán al ser consumidos.

SC jamás es discutido explícitamente, lo que indica que se supone que no es problemático.

En lo que resta de esta sección explicaremos por qué razones el supuesto de correspondencia es insostenible y de qué manera ello hace que la teoría estándar sea ineficaz para contribuir a que la soberanía de que goza el decisor sea ejercida exitosamente.

Preferencias y utilidades son conceptos bien distintos y nada garantiza, a priori, la correspondencia postulada. Aunque en la vida ordinaria el término “preferencia” refiere a creencias acerca de ciertos estados psicológicos de satisfacción, la teoría estándar lo interpreta como si se tratara meramente de una relación *lógica*, de la cual solo interesan sus propiedades. Para clarificar el uso que hacemos aquí de los conceptos de “preferencia” y “utilidad”, distingamos entre dos tipos de estados mentales: las *preferencias* acerca de entidades y las

experiencias de “consumo” de esas entidades. La distinción tiene un aspecto conceptual (epistémico) y un aspecto temporal. En el plano *temporal*, el concepto de preferencia es ex – ante. El de experiencia hedonística es ex – post. Esta distinción está diseñada para dar cuenta de elecciones que, sobre la base de preferencias, preceden en el tiempo a la obtención de resultados y a su consiguiente “consumo”. Toda teoría estándar de la decisión bajo riesgo o incertidumbre adopta una perspectiva ex – ante. En el plano *conceptual*, las experiencias hedonísticas (utilidades) serán consideradas aquí como sensaciones de placer o displacer generadas por el consumo de determinadas entidades, y las preferencias como representaciones (creencias) actuales del grado de satisfacción futura que se obtendrá de dicho “consumo”. En términos de lo expuesto al comienzo de este trabajo, las preferencias sobre resultados no son más que las asignaciones de utilidad que éstos reciben antes de ser obtenidos. Tales asignaciones podrían revelarse luego inapropiadas. Desde nuestro punto de vista lo importante es el *contenido* de las creencias (preferencias), no su estructura lógica.

Valiéndonos de estas precisiones la relación entre preferencias y utilidades asumida en SC puede ser expresada ahora de la siguiente manera. Sea UX la *utilidad* (satisfacción) que proporciona el consumo de X a un sujeto, y $C(UX)$ la *creencia* en que X proporcionará dicha utilidad (satisfacción) en un acto de consumo futuro. En el marco de la teoría estándar, las preferencias poseen las dos siguientes propiedades:

a) son *creencias* (hipótesis) actuales acerca de experiencias de consumo futuras;

b) son creencias *correctas*: es decir, $C(UX) = UX$

El punto central que merece ser destacado es que SC asume la validez de (b): garantiza que las preferencias reflejan adecuadamente las utilidades futuras^{xiv}.

La idea de preferencias como creencias correctas de la satisfacción futura que proporcionarán bienes alternativos disponibles, es importante en economía porque hace plausible la asociación entre preferencias y *racionalidad*. Si hubiera una completa desconexión entre las representaciones actuales y las experiencias posteriores, no podría decirse que

es racional elegir lo más preferido. Porque, recordémoslo, el objetivo de la elección es, si se me permite una licencia, “maximizar” la experiencia resultante de la elección (la utilidad), no la representación (*qua* representación) de las mismas. Si las preferencias sobre bienes no fueran un buen indicador de la satisfacción que se obtendrá de su consumo, no sería fácil de defender la idea de que es racional elegir aquello que se prefiere^{xv}.

A pesar de su ubicuidad, SC es un supuesto problemático. En tanto creencias, las preferencias deberían ser tratadas como se trata a las hipótesis (Boland, 1982). Sin embargo, no se lo hace. Comparemos la elección entre bienes con la elección entre hipótesis científicas. Un conjunto de hipótesis empíricas alternativas no se consideran igualmente correctas; todo lo contrario: si se nos permite abusar una vez más del lenguaje, “maximizar”, en este contexto, podría consistir en elegir la que se cree es una mejor representación de la realidad, elección que no es ajena por completo a consideraciones empíricas (ex – post). Las preferencias, en cambio, son consideradas a la par en cuanto a su verdad: se asume que reflejan adecuadamente la satisfacción que proporcionará cada bien al individuo. Como las preferencias son un tipo de creencias

que no necesitan ser contrastadas, la satisfacción que se obtendrá ex – post del acto de consumo de cada uno de los resultados en caso de ser obtenidos, que es el objetivo real y único de la elección ex – ante, puede ser conocida de antemano examinando el ranking de preferencias. ¿Cuán sustentable es este supuesto?

Supongamos un mundo en que los sujetos eligen entre loterías sobre la base de sus preferencias, pero nunca tienen la posibilidad de verificar si sus preferencias a – priori se corresponden con sus experiencias posteriores. La acción racional es maximizar la utilidad *esperada*, ya que no hay manera de acceder a la utilidad *experimentada*. Las preferencias, en cuanto a representaciones de las experiencias (utilidades), no pueden ser testadas. Es el mundo platónico de la caverna. Pensemos ahora en una situación más realista en que los individuos tienen la chance de acceder a los resultados de sus elecciones, lo que les permite contrastar sus creencias previas. Supongamos que descubren que las utilidades anticipadas (preferencias) de los bienes que componen las loterías no coinciden con las utilidades que efectivamente les brindan una vez obtenidos (consumidos). Y que esta falta de correlación es extendida y

sistemática. ¿Es racional en estas condiciones maximizar la utilidad *esperada*? Consideremos dos situaciones:

Caso 1. Si la brecha entre preferencias y utilidad experimentada permanece inalterada, naturalmente que no lo es, ya que lo que realmente interesa no es maximizar las estimaciones ex – ante, sino las experiencias ex – post, y no hay porque asumir una relación de correspondencia entre ambas. Dicho en otros términos, no hay razón para aceptar SC.

Caso 2. Los sujetos aprenden y progresivamente ajustan sus preferencias a sus experiencias. Puede asumirse que, transcurrido un lapso de tiempo, el ajuste es perfecto y entonces las preferencias se corresponden con las experiencias. Este es precisamente el punto de partida de la teoría estándar. En estas condiciones, el cálculo de utilidad *esperada* es un buen indicador de la utilidad que el individuo obtendrá posteriormente, una vez develado el resultado de su elección. En estas condiciones, la teoría estándar de la decisión adquiere relevancia.

Sin embargo, esta situación de ajuste entre creencias ex – ante y vivencias ex – post, favorable para la teoría de la decisión estándar, supone condiciones que la reflexión y la experiencia empírica sugieren inobtenibles. Por un lado, suponiendo que los individuos son capaces de aprender (incorporar las experiencias pasadas), ello funcionaría asumiendo un mundo estático (los gustos no cambian y no surgen nuevos bienes, de los cuales no se tenga experiencia de consumo). Esta es otra manera de decir que el aprendizaje (en su totalidad) ya ha tenido lugar. Pero en un mundo cambiante, la situación es diferente. Si surgen nuevos objetos, no experimentados (consumidos) aún, las “preferencias” de los individuos respecto de ellos no tiene una base firme. Sólo una vez que se los ha experimentado las preferencias se vuelven representativas. En segundo lugar, numerosos experimentos psicológicos muestran que los individuos no aprenden en el sentido que se necesitaría para que el supuesto de correspondencia fuera adecuado, ni siquiera en un mundo repetitivo y relativamente poco complejo^{xvi}.

III. Función normativa *objetiva* de la teoría económica de la decisión

En la sección anterior cuestionamos la capacidad de la teoría estándar de la decisión para contribuir a que el sujeto pueda realizar los propósitos que *subjetivamente* persigue. Sin embargo, aunque la teoría estándar asume (a-críticamente) que es satisfactoria desde el punto de vista subjetivo, pretende ser algo más que eso. No se contenta con ser un mero instrumento que se pliega sumisamente a los propósitos de los agentes, sino que pone un límite *objetivo* a la soberanía del decisor, permitiéndole rectificar sus decisiones en ciertas circunstancias. Desde esta perspectiva, la valía de sus normas reside fundamentalmente en las ventajas *objetivas* que obtiene de ellas quien recibe sus instrucciones. Estas ventajas pueden ser de dos tipos, según afecten a la *forma* de las preferencias (consistencia) o a sus *contenidos*.

Un aporte normativo innegable de la teoría y, sin duda, el más reconocido, consiste en inducir *consistencia* en las elecciones de los agentes. Como vimos, según la teoría normativa estándar de la decisión los individuos son racionales en tanto sus preferencias son consistentes. La exigencia de consistencia parece ser objetivamente razonable. En esto se basa precisamente la posibilidad de eludir el mecanismo de la “bomba de dinero”: quien viola el axioma de transitividad,

puede que obtenga lo que persigue, pero, a la vez, se expone a ser explotado por quien respeta dicho axioma. Aunque este argumento es convincente, considerado a-priori, puede plantearse una objeción práctica a la exigencia de consistencia: existe evidencia empírica que sugiere que a menudo los agentes violan este requisito, lo que ha conducido a modificaciones en la teoría estándar que prescinden del axioma de transitividad. Por otra parte, la exigencia de consistencia no se reduce a tener preferencias transitivas, sino que va mucho más allá: debe contarse con preferencias *bien formadas*. Pero, como hemos mostrado en la sección I, la evidencia que sugiere que los individuos no poseen preferencias de esta naturaleza es considerable y robusta. La norma que exige consistencia *no sería factible* en el dominio de los humanos. Y si no lo es, no puede ser una buena instrucción para los individuos.

Pero no es este el punto que deseamos enfatizar en esta sección. Lo que deseamos subrayar es que aunque el desiderata de consistencia de la teoría estándar pudiera ser alcanzado por los seres humanos, ello aún resultaría *insuficiente* para asegurar la racionalidad del comportamiento, dado que la consistencia con los axiomas de EUT no impide

que el sujeto sea víctima de la bomba de dinero. Supongamos que el sujeto S1 prefiere \$X a \$Y, y prefiere \$Y a \$Z. En tal caso, es racional cambiar \$Z por \$Y, en una primera ronda, y luego cambiar \$Y por \$X. Supongamos ahora que $X = 0$, $Y = 3$ y $Z = 10$. S1 comenzaría la ronda de intercambios entregando \$10 para obtener \$3 y terminaría con nada. Veamos otro ejemplo. Un sujeto S2 prefiere monedas grandes a monedas más pequeñas. Pero el valor monetario de las mismas es inverso a su tamaño. También en este caso sería desplumado, sin necesidad de que su comportamiento sea inconsistente^{xvii}. Se imponen dos conclusiones.

1) para que la pérdida de recursos tenga lugar, no es necesario poseer preferencias inconsistentes. Puede ser suficiente con tener “malas” preferencias (malas, en un sentido *sustantivo* que será aclarado más adelante). Un sujeto con preferencias “bien formadas” puede iniciar intercambios que lo dejen con menos recursos que los poseídos al inicio.

2) La prescripción “Maximice su utilidad” no es normativamente defendible en estos casos. A sujetos como

S1 y S2 debería decirseles, más bien, “revise sus preferencias” (no la estructura de las mismas, que es la indicada por la teoría y no objetamos, sino su *contenido*).

Que las elecciones de S1 y S2 carezcan de racionalidad *económica* presupone que existe un ranking superior al de ambos sujetos, en un sentido *objetivo* del término. Que el individuo tenga un conjunto de preferencias consistentes y maximice su utilidad es insuficiente para que su ordenamiento individual coincida con el objetivamente superior. Es más, no hay garantía de ello aún en el caso en que experimente ex – post lo que esperaba ex – ante (es decir, aún en el caso de que se cumpla SC). Uno puede actuar consistentemente, quedar subjetivamente satisfecho con las elecciones realizadas y terminar en peores condiciones que al inicio en términos de recursos. La racionalidad *económica* impone restricciones objetivas adicionales a la mera coherencia y a la soberanía decisional de los agentes. Para que las elecciones sean racionales en un sentido económico significativo, deben estar restringidas por la siguiente regla.

Regla {R}: Los sujetos *deben* preferir más a menos bienes.

{R}vale en el caso en que los objetos de elección son conmensurables entre sí o expresables en valores monetarios. En aquellos casos en que los bienes no son valores monetarios, o no son homogéneos (por ejemplo, la elección entre una pera o una manzana) la regla no puede ser aplicada y la racionalidad se reduce a consistencia y la maximización de las preferencias (subjetivas) del agente. Pero en cuanto la elección es sobre un conjunto de opciones homogéneo o conmensurable, del tipo de elegir entre valores monetarios diferentes o entre dos manzanas y dos manzanas y una banana, la regla entra en vigencia.

Esta regla comanda a SC: aunque los sujetos obtengan lo que subjetivamente desean, si su elección es objetivamente inadecuada lo lamentarán (algunos dirían algo todavía más fuerte: terminarán siendo expulsados del mercado). Pareciera, pues, que parte del valor normativo objetivo de la teoría de la decisión económicamente significativa proviene de una instrucción que es *externa* a la teoría estándar de la decisión.

Conclusiones

En su interpretación normativa, la teoría estándar de la decisión asume que las instrucciones que brinda a los agentes son *valiosas para quienes las ponen en práctica*. He aquí una cierta ambigüedad, que refleja dos concepciones diferentes, potencialmente conflictivas, del servicio que la teoría debe prestar a los individuos. De una parte, ligada a una larga tradición en economía que piensa a la ciencia económica como un saber superior al de los propios actores, el papel normativo de la teoría consiste en identificar inconsistencias en el patrón de elección de los individuos y sugerir *rectificaciones* que *mejoren* objetivamente sus elecciones. De otra parte, por estar ligada estrechamente a la doctrina individualista, instruye qué deben hacer los agentes para alcanzar los *propósitos que ellos mismos desean obtener*. Esto permite distinguir en la interpretación normativa de EUT dos papeles diferentes: uno *objetivo* y otro *subjetivo*.

El núcleo normativo de la teoría estándar de la decisión puede ser sintetizado en la máxima "Maximice su utilidad anticipada". Tal instrucción resulta objetiva y subjetivamente

defendible a condición de que se cumplan las siguientes condiciones:

1) Los individuos pueden aprender a comportarse consistentemente (EUT induce consistencia en sus elecciones)

2) Se verifica el Supuesto de Correspondencia (SC) entre preferencias y utilidades;

3) Se cumple la regla que sostiene que los individuos deben preferir más a menos bienes.

(1) y (3) son necesarios para la eficacia normativa *objetiva* de EUT (mejoran y, eventualmente, corrigen sus decisiones), en tanto que (2) es necesario para garantizar su *relevancia subjetiva*. Sin embargo, tales requerimientos son problemáticos. Un primer conjunto de resultados empíricos robustos (constituidos por una diversidad de “efectos”, entre los que se destacan los “framing”, y la acreditada incapacidad de los individuos para aprender en una amplia variedad de situaciones) muestra que (1) no se cumple. Asimismo, un segundo conjunto no menos robusto de resultados experimentales (entre los que se destaca la incapacidad de

los individuos para anticipar la utilidad que obtendrán de los bienes elegidos al momento de consumirlos), muestra que (2) también es violado por los seres humanos. Finalmente, aunque (3) parece ser factible en el dominio de los individuos, se trata de una regla *externa* (de carácter ad – hoc) a la teoría estándar de la decisión, que debe ser adjuntada para evitar que la maximización de la utilidad, en el sentido estándar del término, conduzca a elecciones económicamente irracionales.

BIBLIOGRAFÍA

Arrow, K., Colomatto, E., Perlman, M., and Schmidt, M., (eds), (1999), *The rational Foundations of Economic*

Alchian, Almer, (1953), "The Meaning of Utility Measurement", *The American Economic Review*, Vol. 43, No. 1, pp. 26-50.

Behaviour, Proceedings of the IEA Conference held in Turin, Italy, Great Britain, MACMILLAN PRESS LTD.

Bernoulli, D. (1738), Exposition of A New Theory on the Measurement of Risk, *Econometrica*, 22(1), pp. 23 – 36 (El trabajo original fue publicado en 1738).

Ellsberg, D., (1961), "Risk, Ambiguity and The Savage Axioms", *The Quarterly Journal of Economics*, Vol. 75, N° 4, pp. 643 – 669.

Friedman, M., (1955), "What All is Utility?", *The Economic Journal*, N° 259, Vol. LXV, pp. 405 – 409.

Friedman, M. y Savage, L. J., (1948), "The Utility Analysis of Choices Involving Risk", *Journal of Political Economy*, LVI, pp. 279 – 304.

Friedman, M. y Savage, L. J., (1952), "The Expected utility Hypothesis and the Measurability of utility" (*The Journal of Political Economy*, Vol. LX, N° 6).

Hausman, D., (1992), *The Inexact and Separate Science of Economics*, Cambridge University press, Cambridge.

Kahneman y Tversky. (1979). Prospect theory: An analysis of decisions under risk. *Econometrica*, 47, 313-327.

Kahneman, D., and Thaler, R., 1991, "Economic Analysis and the Psychology of Utility: Applications to compensation Policy", *The American economic Review*, Vol. 81, N° 2, pp. 341 – 346.

Kahneman, D., (2003a), "Maps of Bounded Rationality", *The American economic Review*, Vol. 93, N° 5, pp. 1449 - 1475.

Kahneman, D., (2003b), "A psychological Perspective on Economics", *AEA*, 2003, p. 162).

Kahneman, D, and Sugden, R., (2005) "Experienced Utility as a Standard of Policy Evaluation", *Environmental & Resource Economics* (2005), 32: 161 - 181.

Kahneman, D. and Thaler, R., (2006), "Utility Maximization and Experienced Utility", *Journal of Economic Perspectives*, Volume 20, N° 1, pp. 221 – 234.

Loomes, G. and Sugden, R., (1982), "Regret Theory: An Alternative Theory of Rational

Choice Under Uncertainty", *Economic Journal*, 92, 805 – 824.

Machina, M., (1982), "'Expected Utility' Analysis Without the Independence Axiom", *Econometrica*, Volume 50, N° 2.

Schoemaker, P.J., "The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations", *JEL*, June 1982.

Slovic, P. and Lichtenstein, S., (1983), "Preference Reversals: A Broader Perspective", *The American Economic Review*, Vol. 73, N° 4, pp. 596 – 605.

Starmer, Chris, (2000), "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk", *Journal of Economic Literature*, Vol., XXXVIII, pp. 332-382.

Tversky, A., (1999), "Rational Theory and Constructive Choice", en Arrow, Colombatto, Perlman and Schmidt, (eds), *óp. cit.*

Tversky, A. and Kahneman, D., (1981), "The Framing of Decisions and the Psychology of Choice", *Science*, Vol. 211, N° 4481, pp. 453-458.

Tversky, A. and Kahneman, D., (1986), "Rational Choice and the Framing of Decisions", *The Journal of Business*, Vo, 59, N° 4, pt. 2.

¹ Más formalmente, si, para simplificar, nos limitamos a loterías de sólo dos resultados, pueden ser representadas de esta manera: $[x, p; y, (1-p)]$, en donde x e y representan los resultados o premios que pueden ser obtenidos si se las elige, y p y $(1 - p)$ las probabilidades asociadas a cada uno de ellos (en nuestro ejemplo, x podría ser la heladera y el microondas, y sería cero premio, y p tendría un valor de 0.2).

¹ Véase Bernoulli, 1954.

¹ Frieman y Savage, 1952.

¹ Alchian (1953) ofrece una presentación sumamente accesible de este procedimiento. Véase también Friedman y Savage, 1948, p. 292.

¹ Los fallidos más conocidos son las paradojas de Allais ("common" y "consequence effects") y Ellsberg (juicios probabilísticos inconsistentes), las "preference reversals" (Slovic y Lichtenstein), y los "framing effects" (Kahneman y Tversky).

¹ Machina (1982),

¹ Loomes y Sugden (1982).

¹ Las principales alternativas introducen transformaciones apropiadas en alguna de las variables clave de EUH, generando "fanning out", al operar sobre las utilidades

(Machina, 1982) y transformando las probabilidades en “decision weights” (Kahneman y Tversky, 1979). Para una descripción de los cambios operados en EUH véase Schoemaker (1982) y Starmer (2000).

¹ Starmer (2000, pp. 360-363), agrupa a las desviaciones persistentes en tres ítems: a) violaciones de la monotonicidad; b) event – splitting effects y c) violaciones de la transitividad.

¹ “The basic principle of framing is the passive acceptance of the formulation given. Because of this passivity, people fail to construct a canonical representation for all extensionally equivalent descriptions of a state of affairs (...). they do not spontaneously transform the representations of puzzles or decisions problems. Obviously, no one is able to recognize ‘137 x 24’ and ‘3288’ as ‘the same’ number without going through some elaborate computations” (Kahneman (2003a, p. 1459).

¹ “Procedure invariance plays an essential role in measurement theories. For example, the ordering of objects with respect to mass can be established either by placing each object separately on a scale, or by placing both objects on the two sides of a pan balance; the two procedures yield the same ordering, within the limit of measurement errors. Analogously, the classical theory of preference assumes that each individual has a well defined preference order (or a utility function) that can be elicited either by offering a choice between options, or by observing their reservation price. Procedure invariance provides a test for the existence of a measurable attribute. It would have been difficult to attribute mass to objects if the orderings of these objects with respect to mass were dependent on the measuring device. *Similarly, it is difficult to defend the proposition that a person has a well-defined preference order (or equivalently a utility function) if different methods of elicitation give rise to different choices*” (Tversky, 1999, p. 189; my italics).

¹ Friedman, 1955.

¹ Para una defensa de la posibilidad de medir la utilidad en el sentido hedonista del término, véase Kahneman y Sugden, 2005.

¹ Si en la consideración de un sujeto S, A es preferida a B, al elegir A el objetivo del sujeto es obtener del consumo ex – post de A la utilidad (experiencia hedonística) que esperaba ex – ante (o, al menos, mayor utilidad que la que obtendría de consumir B). El sujeto tiene una idea (una expectativa) de la utilidad que obtendrá de ese y otros bienes alternativos y es esta la que determina su elección. Esto significa que el sujeto prefiere A a B porque tiene una creencia (hipótesis) acerca del grado de satisfacción que le proporcionarán ambos bienes y puede compararlos a este respecto; y porque cree, además, que el consumo de A le brindará mayor satisfacción que el de B.

¹ Una consecuencia de lo dicho, que no desarrollaremos en este trabajo, es que el intervencionismo, al menos en su forma débil de “paternalismo”, no sería fácilmente deseable en tales condiciones. El supuesto sustantivo de correspondencia es también importante, pues, porque se conecta con la defensa del no intervencionismo (laisse faire).

¹ Al respecto véase Kahneman y Sugden (1985) y Kahneman y Thaler (1986).

¹ La irracionalidad de estos intercambios no depende de que el sujeto pierda todo su dinero; basta con que emprenda una serie de transacciones que culmine con una disminución de su riqueza inicial.

SOBRE LA INUTILIDAD DE LAS ENCUESTAS

Diego Weisman

CIECE-FCE-UBA/CONICET

La idea de que las encuestas no suelen brindar información confiable está bastante extendida, no sólo entre quienes las diseñan y utilizan con regularidad sino entre el público en general. Si bien las metodologías con base estadística intentan minimizar los sesgos, y se confía en la probidad de los científicos sociales para descontar manipulaciones, los resultados muchas veces son desalentadoramente paradójicos.

Por qué fallan las encuestas? Más allá de la honestidad científica de quienes diseñan/aplican las encuestas, o de la implementación de mejoras estadísticas, hay una razón teórica que explica –y predice- la falla de las encuestas, al menos en un dominio relevante. Si esto es así ya no son necesarias las explicaciones *ad hoc* destinadas a digerir el fracaso sin menoscabar la teoría implícita en el uso de encuestas. En otras palabras, el culpable de la falla en las

encuestas puede estar en el supuesto básico que guían su uso: la existencia de preferencias libres y autónomas.

Prospect Theory, la Teoría de la Decisión bajo riesgo propuesta por Kahneman y Tversky en 1979 introduce cambios notables en cuanto a los mecanismos que guían la toma de decisiones. Las variables causales de la elección son, además de los resultados posibles y su probabilidad – únicas consideradas en la teoría clásica-, la *descripción* de la situación planteada. Como las encuestas enfrentan a los individuos con un problema de elección sin tomar en cuenta que la descripción es una de las variables causales de la elección, no sorprende que existan resultados anómalos. Prospect Theory predice dichos anomalías, pero las consecuencias de ampliar el rango de variables en la elección son difíciles de aceptar.

Puede repararse el error? La segunda parte del trabajo presenta una cuestión ontológica que permite abrigar dudas al respecto. La *existencia* de una propiedad comparativa (cómo “más preferido que”, o “más pesado que”) exige la invariancia procedural: dicho de manera escueta, la escala (comparativa) construida a partir de dicha propiedad debe ser

única, e independiente de procedimiento de construcción. En la medida en que esto no ocurra, y que encuestas equivalentes den lugar a escalas de preferencias distintas, las preferencias que se pretende sacar a la luz simplemente no existen.

1-VARIABLES INTERNAS Y EXTERNAS

La economía modela las acciones humanas en contextos de incertidumbre como motivadas por creencias y deseos. Las creencias son probabilísticas, y los deseos se expresan como preferencias sobre resultados posibles.

Cuando el resultado de una acción es incierto, como sucede por ejemplo en una apuesta de ruleta, una regla de decisión posible es la de ponderar el *monto* de ese resultado por la *probabilidad* de su ocurrencia. La idea detrás de esto es razonable: para saber si es conveniente apostar en un juego de azar es necesario saber no solamente cuánto se puede ganar, sino qué probabilidades hay de hacerlo.

La mención a las apuestas no es casual. La teoría de la decisión bajo incertidumbre tiene vínculos históricos y conceptuales estrechos con el estudio de los juegos de azar

emprendido hacia fines del siglo XVII. Por supuesto, allí había premios monetarios, lo que permitía definir la cantidad razonable a pagar por intervenir en un juego cualquiera, donde los premios sean X y las probabilidades asociadas, P , mediante la siguiente operatoria:

TEORÍA DE LOS JUEGOS

$VE(A): \sum X_i \cdot P_i$ [*Variables explicativas: Resultados y Probabilidades*]

Como las probabilidades y los resultados son variables objetivas del juego, en el sentido de que son las mismas para todos los eventuales jugadores, se puede definir algo así como una apuesta racional que es idéntica para todos, al menos en ciertas circunstancias. Piense en la siguiente decisión: pagaría 5 \$ por el juego consistente en tirar una moneda y que le paguen 6 \$ si sale cara, y nada si sale ceca? La respuesta intuitiva es que no, y la teoría de los juegos de azar permite además calcular cuál es la cantidad “justa” a pagar por entrar en ese juego.

La operatoria en la Teoría de la Decisión Racional es ligeramente diferente. El determinante en juego aquí es la

utilidad de los resultados, y no los resultados en sí mismos. Bernouille en 1734 es quien introduce el término utilidad en la ecuación, con el objeto de resolver la célebre Paradoja de San Petersburgo. En la práctica, la modificación de Bernouille permite trascender los límites de los resultados monetarios y expandir así el campo de la teoría de la decisión para *cualquier* resultado, sea en dinero o en bienes sin precio de mercado. En una lotería que en lugar de premios monetarios tenga premios simbólicos, siempre se puede asignarle alguna utilidad a esos bienes, lo cual permite aplicar la TDR en prácticamente cualquier ámbito donde haya riesgo.

TEORIA DE LA DECISION RACIONAL

UE: $\sum U(X_i) P_i$ [Variables explicativas: Utilidad de los resultados y Probabilidades]

Nótese que a diferencia de la Teoría de los Juegos, TDR contempla la posibilidad de que dos individuos tomen decisiones racionales diferentes, básicamente porque una de las variables –la Utilidad asignada a X- es subjetiva. En la mayoría de los casos hay una respuesta racional unívoca a la pregunta de si es razonable apostar y cuánto por un juego de

azar determinado, o alternativamente la teoría de los juegos de azar podría dar una predicción determinada respecto a la elección de un jugador racional enfrentado a un par de opciones (del tipo “elija entre 1- apostar a cara en una tirada de una moneda y ganar 10 \$ si acierta, y nada si no acierta o 2-tener 4 \$ con certeza”) Nada de esto ocurre cuando con la TDR. Como aquí es posible asignarle al mismo premio (10\$ o 4\$) distintas utilidades, la elección de una opción sobre otra permanece indeterminada. Lo que puede testarse acá es la *consistencia* que se revela en la elección entre opciones, y a esto se llama “decisión racional”.

Ahora bien, una cuestión interesante es que la consistencia puede testarse empíricamente, enfrentando a sujetos con decisiones riesgosas alternativas (loterías) y viendo de qué manera eligen. En la literatura económica son bien conocidas las anomalías y paradojas a las que el programa condujo, como las de Allais, Ellsberg, et al. (cfr. Hausman, 1992)

Así como TDR resuelve la paradoja de San Petersburgo mediante el expediente de reemplazar los premios monetarios por la utilidad de los mismos, *Prospect Theory*, propuesta por Kahneman y Tversky en 1979, resuelve otras

paradojas, como las *preference reversals*, mediante una modificación todavía más radical. A los fines que se persiguen, conviene expresar las variables causales de la decisión como las siguientes.

PROSPECT THEORY

$UP: \sum U^*(X_i) \in (P_i)$ [*Variables explicativas: Resultados, Probabilidades y Descripción*]

La ruptura sobre la que quiero llamar la atención es la siguiente: a diferencia de las variables presentes en la Teoría de los Juegos y en la Teoría de la Decisión Racional, *Prospect Theory* introduce una variable que es de *naturaleza* diferente, en la medida en que ya no depende de la situación “objetiva” –el juego o la lotería- sino de una consideración externa: *la descripción de la situación objetiva*. Voy a llamar a esta tercer variable una variable *externa*, por razones obvias.

2-VARIABLES CAUSALES DE LA DECISIÓN

Las encuestas pretenden sacar a la luz preferencias ofreciendo a los individuos opciones para que decidan entre

ellas. Qué significa *decidir entre opciones*? La teoría de la decisión responde a esa pregunta, pero antes de volver a ella veamos cuáles son los presupuestos básicos de las encuestas:

1-Existe una escala de preferencias que es independiente de y previa a la encuesta. La encuesta en sí es un mero instrumento para determinar una entidad ajena. Llamaremos a esta característica supuesta la de “preferencias libres”

2-La decisión entre las opciones ofrecidas está motivada únicamente por características internas de esas opciones. El objeto de preferencia es justamente un objeto, y son sus rasgos intrínsecos (y quizás los deseos subjetivos del agente) los que explican la elección. Llamaremos a este segundo supuesto el de “preferencias autónomas”, porque salvo excepciones las decisiones aparecen motivadas por la relación de un sujeto con el mundo, sin posibilidad de que un tercero interfiera de manera sistemática en esta relación.

Podemos afirmar ahora que las encuestas pretenden idealmente sacar a la luz preferencias libres y autónomas.

Ahora bien, si *Prospect Theory* puede predecir la conducta individual es porque ha identificado una variable causal relevante en las decisiones humanas: además de los resultados y las probabilidades, la *descripción* de la situación determina las decisiones de los agentes.

Las implicaciones de este hecho –la incorporación de una variable externa como una variable causal relevante en la decisión humana– son difíciles de asimilar, porque requieren afrontar la compleja tarea de desmontar una antigua metafísica construida sobre la decisión humana que sin duda excede el campo de la Teoría de la Decisión. Sin embargo, parece cierto que una primera consecuencia es la siguiente:

A diferencia de la concepción tradicional de predicción, en la cual el teórico se comporta como un verdadero espectador frente al objeto de estudio, en las predicciones de la decisión humana es necesario intervenir para asegurar el éxito predictivo.

La intervención del teórico es trivialmente cierta en contextos de laboratorio. En el caso de la decisión humana, la situación de laboratorio contempla enfrentar al sujeto con opciones

(loterías) que se pueden ver como una tríada de resultados, probabilidades asociadas y una cierta descripción. Un ejemplo es el ya clásico de la Enfermedad Asiática:

*The Asian disease*¹⁶

Imagine that the United States is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed.

Assume that the exact scientific estimates of the consequences of the programs are as follows:

If Program A is adopted, 200 people will be saved

If Program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no people will be saved

¹⁶ Cfr. Kahneman y Tversky, *Maps of Bounded Rationality* (2003) pp.1458.

La mayoría de los sujetos experimentales elijen el Programa A sobre el B, revelando aversión al riesgo. Veamos el siguiente par de alternativas:

If Program A' is adopted, 400 people will die

If Program B' is adopted, there is a one third probability that nobody will die and a two-thirds probability that 600 people will die

Ahora la mayoría de los sujetos eligen el Programa B sobre el A, a pesar de que ambos pares de opciones describen el *mismo* conjunto en términos de vidas perdidas y salvadas. ¿Qué significa esto?

Prospect Theory predice la elección de A en el primer caso y la elección de B en el segundo, porque el teórico tiene en sus manos las tres variables causales de la elección. Conoce los resultados, las probabilidades asociadas y la *descripción* de la situación que enfrenta el sujeto (en el primer caso, el problema se plantea en términos de vidas salvadas, en el segundo de vidas perdidas).

Pero qué sucede en contextos abiertos? Es posible predecir la conducta de los individuos cuando se sale de las condiciones restrictivas del laboratorio? Esto nos lleva a la segunda consideración:

Los resultados experimentales conseguidos en laboratorio no son directamente aplicables para predecir la conducta en contextos abiertos, porque una de las variables causales determinante de los resultados (i.e., la descripción) no puede conocerse de antemano.

Para ilustrar el caso, veamos el siguiente ejemplo:

- *Ejemplo del apostador:*

Consider the following cases: at the end of a day an individual who has lost \$140 has to decide whether to pay \$10 for a prospect that may pay \$150 with probability 1/15 or nothing with probability 14/15 (Tversky and Kahneman, 1981, p. 456). Clearly his options are to bet or not to, but these two alternatives may be described in many different ways. For instance, he may take into account his accumulated loss of \$140 or he may leave the past performance out of

consideration and start anew. The prospects he is going to consider will depend on where his reference point is located.

- *Case 1. The individual does not take into account past losses. His reference point is in this case \$0, and his options are the following:*

A: Not to Bet (A) = 0

B: To Bet (B) = (-\$10, 14/15; +\$140, 1/15)

- *Case 2. The individual considers past losses. His reference point is now -\$140, and his options change accordingly:*

C: Not to Bet C: (-\$140)

D: To Bet D: (-\$150, 14/15; \$0, 1/15)

Como es posible predecir si el apostador va a apostar o no?
La respuesta es que *no puede saberse de antemano*.¹⁷ No

¹⁷ Las consecuencias de los *framing effects* están desarrolladas en Marqués-Weisman (2009)

puede hacérselo, una vez más, porque su decisión está causada en parte por una variable causal que no depende de las circunstancias “objetivas” de la situación –resultados o premios y probabilidades-, sino que depende de la *descripción* que el sujeto asuma como significativa. ¿Considerará el apostador sus pérdidas pasadas? En tal caso, va a apostar. ¿Verá cada apuesta como un evento independiente? En ese caso, no va a apostar. No hay manera, sin intervención directa del teórico, en que pueda conocerse de antemano -predecirse- la elección individual. La predicción solamente puede ser exitosa en laboratorio, y si puede extenderse esa capacidad predictiva a la realidad es en la medida en que se tenga algún poder sobre las descripciones que enfrenta el sujeto.

La relación con el problema de las encuestas puede ahora establecerse de manera explícita. Tanto en condiciones de laboratorio como en contextos abiertos, la manipulación de las descripciones se muestra necesaria para predecir/producir las elecciones. Las encuestas presentan opciones *bajo una descripción*, al tiempo que pretenden elucidar preferencias que son *previas* a esa descripción. En la medida en que *Prospect Theory* esté en lo correcto, los encuestadores están

produciendo (i.e., manipulando) las preferencias en el acto mismo de elucidarlas.

3-El mito de las preferencias libres y autónomas

En mi experiencia, cada vez que el argumento anterior es presentado al público, la reacción de una parte considerable del auditorio es pensar algo como lo siguiente: “Bueno, probablemente *algunos* sujetos puedan ser manipulados en sus elecciones, por inadvertencia o falta de educación. Es posible que se pueda engañar a la gente con respecto a qué es lo que verdaderamente prefiere. Pero en la medida en que se los eduque o aprendan por sus propios medios, sus verdaderas preferencias van a reaparecer”

Esta no es una idea extraña. Remite a los oscuros orígenes históricos en los que fue forjado el concepto de un hombre autónomo y racional –un sujeto iluminista, portador de preferencias que se revelaban en, pero que eran previas a sus elecciones. También está implícita en la justificación conceptual de la democracia, y de los sistemas económicos: Si la elección ha de ser libre (la elección democrática, o la elección de régimen de vida, o la elección de bienes y

servicios) es porque detrás de ella hay preferencias autónomas, no manipuladas por terceros. Justamente era la clase de preferencias libres presupuestas por la teoría de la decisión tradicional (EUT), pero que se pierden irremisiblemente con *Prospect Theory*.

Volvamos a los supuestos subyacentes al uso de encuestas. Estos eran:

1-Existe una escala de preferencias que es independiente de y previa a la encuesta; es decir, las preferencias son “preferencias *libres*”

2-Las preferencias además son *autónomas*, en el sentido de que no intervienen terceros.

Es fácil ver ahora que el supuesto 1 no se cumple, porque las encuestas incorporan necesariamente una determinada *descripción* de la situación, y la escala de preferencias no es independiente a ella. El supuesto 2 cae porque la descripción puede ser manipulada; más aun, si pueden predecirse las elecciones es porque pueden ser manipuladas en un sentido definido.

Podemos ahora responder a la pregunta de por qué fallan las (al menos, algunas) encuestas. En la medida en que las encuestas pretenden extraer información referente a preferencias libres y autónomas, están tratando de elucidar una entidad mítica. De manera semejante a como los químicos anteriores a Lavoisier andaban a la caza del flogisto, los encuestadores persiguen una quimera con instrumentos particularmente inadecuados para ello. Puede repararse este error?

2-

Muchos economistas piensan con alguna razón que su disciplina ha descuidado la base empírica. Embarcados en un honesto intento por restaurar la conexión perdida con la experiencia, intentan implantar “métodos empíricos”, entre los cuales están los de encuestas. Y muchas de estas encuestas intentan averiguar cuáles son las preferencias de los individuos, mediante el expediente de entrevistar personas y darle opciones a elegir. Cada persona tiene una escala de preferencias, se piensa, y esta escala se puede sacar a la luz observando sus elecciones (o preguntándole de manera directa qué elección tomaría)

En este panorama, una condición que se da por supuesta es que orden comparativo de preferencias es independiente del procedimiento utilizado para elucidar dicho orden. Llamaremos a este supuesto “invariancia procedural” Tversky lo define y señala su importancia de la siguiente manera:

“Procedure invariance plays an essential role in measurement theories. For example, the ordering of objects with respect to mass can be established either by placing each object separately on a scale, or by placing both objects on the two sides of a pan balance; the two procedures yield the same ordering, within the limit of measurement errors. Analogously, the classical theory of preference assumes that each individual has a well defined preference order (or a utility function) that can be elicited either by offering a choice between options, or by observing their reservation price. Procedure invariance provides a test for the existence of a measurable attribute. It would have been difficult to attribute mass to objects if the orderings of these objects with respect to mass were dependent on the measuring device. Similarly, it is difficult to defend the proposition that a person has a well-defined preference order (or equivalently a utility function) if

different methods of elicitation give rise to different choices”
(Tversky, 1999, p. 189).

Por qué es tan importante contar con una orden único, con una escala que sea independiente del método de medición? Por la siguiente razón: una condición necesaria para poder afirmar que una propiedad *existe* es que dos métodos de medición equivalentes arrojen como resultado órdenes idénticos. Por ejemplo, establecida una propiedad cualquiera, digamos el volumen, en un cierto dominio –el de los seres vivos-, se puede formar un orden que es único. La escala ordinal del volumen de un elefante, un ruiñeñor y una hormiga es *la misma*, ya sea que se la construya mediante métodos gravimétricos o geométricos. Esta circunstancia nos permite afirmar que el volumen es una propiedad que existe independientemente de los métodos que se usen para determinarla, lo cual es una forma trabajosa de decir que cualquier método adecuado arrojará el mismo orden comparativo (volumen del elefante > volumen del ruiñeñor > volumen de la hormiga) *Mutatis mutandis*, si alguien quisiera convencernos de que *existe* una propiedad comparativa que es, digamos, el color del aura de los animales (el cual se nos informa que puede ser más claro o más oscuro), se verá a sí

mismo en un aprieto si diversas mediciones dieran lugar a escalas ordenadas de manera distinta, donde unas veces el aura del elefante fuera más oscura y otras veces más clara que el de una hormiga.

Si este fuera el caso, dos caminos se abren: o bien podríamos decir que los procedimientos de medición no son realmente equivalentes, o bien que la magnitud que se intenta establecer simplemente no existe en ese dominio. Asumiendo que dos métodos sean equivalentes entre sí, ordenar cuerpos por la cantidad de flogisto o el color del aura muy probablemente dará lugar a órdenes diferentes. Esto puede ser percibido al principio como una falla en la medición, pero si la situación se mantiene en el tiempo quizás daría lugar a que se sospeche de la existencia de la entidad misma.

Para resumir lo anterior en forma de pregunta, ¿qué sucedería si dos métodos para medir la propiedad X, del que hay sobradas razones para suponer equivalentes, dieran lugar a órdenes distintos? Creo que el lector estará de acuerdo en afirmar que la respuesta más sobria sería asumir simplemente que la *propiedad X no existe*, al menos en los

dominios donde no rija la invariancia procedural. *Tal es el caso de la propiedad “ser más preferido que” en el dominio de las elecciones humanas.*

4-Esbozo de un planteo empírico

Avancemos un paso más: es posible determinar en qué dominios existe invariancia procedural? Esto sería importante porque permitiría detectar de antemano en qué dominios la propiedad que pretende elicitarse existe, y por ende daría lugar a separar propiedades como el volumen de propiedades como la presencia de flogisto. Para el caso que nos ocupa, puede pensarse un esbozo de planteo empírico muy simple. Consiste simplemente en diseñar dos preguntas, o series de preguntas que sean equivalentes entre sí, y analizar si efectivamente dan lugar al mismo orden en las respuestas. Por ejemplo, se puede preguntar:

A: Qué lo haría más feliz (o qué preferiría), la situación X o la Y?

B: Qué lo haría menos infeliz, la situación X o la Y?

Si se acepta que “más feliz” es equivalente a decir “menos infeliz”, entonces tenemos dos procedimientos A y B que son equivalentes. Para que pase el test de existencia, el orden revelado por las respuestas debería coincidir. Lamentablemente, es posible conjeturar que no será el caso: los órdenes observados serán diferentes.

En efecto, el ejemplo de la felicidad no es arbitrario. Recientes trabajos (Ormerod y Johns, 2007, 2008, 2009) señalan que *“statistically significant correlations between time series happiness data and other important socioeconomic indicators cannot be found, that the nature of happiness scales makes them insensitive and difficult to compare with most other economic data, and that using time series happiness data for policy-making creates several undesirable problems”*, por lo que *“time series data on happiness tells us nothing”*.

No deberían sorprendernos estos resultados. Las series de tiempo de indicadores como el de felicidad percibida se construyen después de todo bajo el supuesto de preferencias libres y autónomas, una entidad cuya existencia es, como hemos visto, altamente dudosa.

5-Sirve estandarizar las encuestas?

Si dos encuestas equivalentes, digamos A y B, dan lugar a órdenes distintos en las respuestas, entonces cabe la posibilidad de que las preferencias que pretenden sacarse a la luz (unas preferencias *previas* a la descripción de las opciones y *autónomas*) sean una suerte de concepto semejante al flogisto para la química pre Lavoisier.

¿Mejora la situación si se elige una de las encuestas, la A o la B, y se mantiene en el tiempo? Es decir, si se recurre a *encuestas estandarizadas*? Esta es una manera extraña de resolver la cuestión, pero es la respuesta habitual que suele recibirse cada vez que se pone en duda la probidad de las encuestas. Observemos primero que mantener una encuesta estandarizada cortaría de raíz el problema: para que se manifieste la dualidad en los órdenes tiene que haber al menos dos encuestas alternativas. Pero es una solución trivial, por supuesto. Para ilustrar esto, vamos a usar dos experimentos mentales, del tipo de la Tierra Gemela.

Supongamos que hubiera dos elementos en el Planeta X –un mundo idéntico a la Tierra-, y que los científicos de ese

planeta quisieran determinar la densidad relativa de dichos elementos. Supongamos también que, por razones misteriosas, cuando la prueba se realiza en los laboratorios del hemisferio occidental del Planeta X, la densidad del elemento 1 fuera mayor a la densidad del elemento 2, pero cuando la misma prueba se realiza en el hemisferio oriental la densidad del elemento 1 fuera menor que la del elemento 2. Como es fácil de imaginar, este absurdo es motivo de gran escándalo en la comunidad científica.

Pasada la sorpresa inicial, y luego de algunas discusiones, se decide con alguna solemnidad que la densidad es válida solamente cuando el procedimiento para medirla se fija en el hemisferio occidental. Esto preserva el orden –y la paz mental de los científicos- y por ridículo o antojadizo que parezca *es el sucedáneo de quienes abogan por las encuestas estandarizadas.*

El problema que queda abierto no es menor: es el de si realmente esas mediciones (consistentes) son una medida de *algo*, o de si existe una variable causal oculta que afecta la densidad relativa de los elementos, no contemplada en la teoría actual.

Volviendo al tema de las encuestas, la variable causal propuesta por *Prospect Theory* –la descripción- obligaría a revisar el concepto de preferencias libres, previas a la descripción, con sus consecuencias difíciles de asumir para la teoría política, la libre elección, las posibilidades de manipulación, etc. Y sin embargo, este no es el peor de los mundos posibles.

En el Planeta Y, los científicos consideran que existe una propiedad –la *dadisned*-y se encuentran con que métodos equivalentes dan lugar a órdenes de *dadisned* diferentes: usando el método A, por ejemplo, se obtiene a veces que la *dadisned* del elemento 1 es mayor que la del elemento 2, y otras veces ocurre lo contrario. A diferencia del Planeta X, aquí no se encuentra nada *sistemático* en la reversión de órdenes de *dadisned*.

De qué valdría, en el Planeta Y, fijar el método A para las mediciones? No es más sencillo decir que esa propiedad tan caprichosa no existe como tal?

El caso de las preferencias libres y autónomas no es tan desesperado como el del Planeta Y. Muchas de las

anomalías que existen son *sistemáticas*, lo que permite abrigar esperanzas de que haya una variable oculta que no se está tomando en consideración, sea la descripción o alguna otra. Quizás, después de todo, las encuestas tengan alguna utilidad, aunque es necesario mucho trabajo teórico futuro para dotarlas de pleno sentido.

Bibliografia

Barberis, N. and Thaler, R., 2003, "A Survey of Behavioral Finance", in G.M Constantinides, M.

Camerer, C., and Loewenstein, G., 2002, "Behavioral Economics: Past, Present and Future" (<http://www.hss.caltech.edu/~camerer/ribe239.pdf>)

Friedman, M, and Savage, L., 1952, "The Expected utility Hypothesis and the Measurability of utility" (The Journal of Political Economy, Vol. LX, N° 6).

Friedman, M., 1955, "What All is Utility", The Economic Journal, Vol. 65, N° 259, pp. 405 – 409.

Grether, D., and Plott, C., (1979), "Economic Theory and the Preference Reversal Phenomenon", American Economic Review 69: 623 – 638.

Hausman, D., (1992), *The Inexact and Separate Science of Economics*, Cambridge University press, Cambridge.

Holt, Ch., (1986), Preference Reversals and the Independence Axiom, *The American Economic Review*, vol. 7, N° 3.

Kahneman, D, (2003a) "Maps of Bounded Rationality: Psychology for Behavioral Economics", *AER*, Vol. 93, N° 5, pp. 1449 – 1475.

Kahneman, D., (2003b), "A psychological Perspective on Economics", *AER*, Vol. 93, N° 2, pp. 162 - 168.

Kahneman, D. and Tversky, A, (1979), "Prospect Theory: An Analysis of Decision Under Risk", *Econometrica* 47, pp. 263-291.

Kahneman, D., Knetsch, J.L. and Thaler, R.H., (1991), "The Endowment Effect, Loss Aversion, and Status Quo Bias", *Journal of Economic Literature*, Volume 5, N° 1, pp. 193 – 206.

Lichtenstein, S., and Slovic, P., (1971), "Reversal Preferences between bids and choicen in gambling decision", *Journal of Experimental Psychology* 89, pp. 46 – 55.

Loomes, G., and Sugden, R., (1982), Regret Theory: An Alternative Theory of Rational Choice Under Uncertainty, *The Economic Journal*, Vol. 92, No. 368 (Dec., 1982), pp. 805-824.

Machina, M., (1982), "Expected Utility' Analysis Without the Independence Axiom", *Econometrica*, Volume 50, N° 2.

Marschack, J., (1950), "Rational Behavior, Uncertain Prospects and Measurable Utility", *Econometrica*, Vol. 18, N° 2, pp. 111 – 141.

Mongin, Ph., (1998), "Expected Utility Theory", in Davis, J., B., Hands, D., W., and Mâki, U., (eds), *The Handbook of Economic Methodology*, Cheltenham, Edward Elgar, pp. 171 - 178.

Schoemaker, P.J., The Expected Utility Model: Its Variants, Purposes, Evidence and Limitations, *JEL*, June 1982.

Slovic, P., and Lichtenstein, S., (1983), Preference Reversals: a Broader Perspective", *The American Economic Review*, Vol. 73, N° 4, pp. 596 – 605.

Starmer, Chris, (2000), "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk", *Journal of Economic Literature*, Vol., XXXVIII, pp. 332-382

Thaler, Richard, 1986, "The Psychology and Economics Conference Handbook: Comments on Simon, on Einhorn and Hogarth, and on Tversky and Kahneman", *The Journal of Business*, Vol 59, N° 4, Part 2, pp. S279 – S284.

Thaler, Richard, 2000, "From Homo Economicus to Homo Sapiens", *Journal of Economic Perspectives*, Volume 14, Number 1, Winter 2000, pp. 133-141).

Tversky, A., (1999), "Rational Theory and Constructive Choice", in Arrow, K., Colombatto, E., Perlman, M., and Schmidt, M., (eds), *The rational Foundations of Economic Behaviour*", Great Britain, MACMILLAN PRESS LTD.

Tversky, A. and Kahneman, D, (1981), "The Framing of Decisions and the Psychology of Choice", *Science, New Series*, Vol. 211, N° 4481, pp. 453 – 458.

Tversky, A. and Kahneman, D. (1986), "Rational Choice and the Framing of Decisions", The Journal of Business, Vol. 59, N° 4, Part 2, pp. S251 – S278.

Von Neumann, J. and Morgenstern, O., 1944, Theory of Games and Economic Behavior; Princeton University Press.

Tercer Sección:
Neurociencia e Interdisciplinariedad

THE POSSIBILITIES OF NEUROECONOMICS:

An account through the lens of economic methodology

Roberta Muramatsu[♦]

ABSTRACT

Neuroeconomics is a recent development within the research agenda called behavioral economics that has become rather popular at the end of the 20th century. The departure point of this paper is the view that the appeal as well as the sources of criticism of neuroeconomics reside on its 'rhetoric and style' that promise to revolutionize economic analysis by employing new brain imaging techniques that uncover heuristic processes and neural mechanisms underlying actual judgment and decision-making and provide accurate direct measurement of economic models' preference and belief

[♦] Assistant Professor of Economics and Member of the Economic Wellbeing and Quality of Life Research Group, Center for Applied Social Sciences, Mackenzie University of São Paulo. Senior Lecturer of Economic History at IBMEC Business School São Paulo, Brazil. Email address: muramatsu@uol.com.br or robertam@isp.edu.br.

parameters. Based on a methodological strategy that does not draw on prepackaged ideas of philosophy of science, we focus on the analysis of actual practice of neuroeconomics so as to assess the logic of emergence, accomplishments and challenges posed to neuroeconomics. We conclude that, despite some rhetorical exaggerations (non-qualified claims) and methodological problems that characterize its practice, the young subfield of neuroeconomics is not a “hype” but genuine hope to contribute to the development of a theoretical approach that is more explanatory in the sense of unearthing processes and mechanisms with major causal roles in production of actual behavior and also more general, in being able to predict phenomena already accommodated by conventional theory as well as those regarded as anomalies.

KEYWORDS: methodology, neuroeconomics, explanation, prediction, experiment

JEL Classification: B40, B41, D11, D91

From a logical point of view therefore, one who aspires to explain or understand human behavior must be, not finally but first of all, an epistemologist.

Frank Knight, Economic Psychology and the Value Problem

The goal of neuroeconomics is to ground economic theory in details of how the brain works in decision making, strategic thinking and exchange. One way to achieve this is to observe processes and constructs which are typically considered unobservable (...)

Colin Camerer, Neuroeconomics: using neuroscience to make economic predictions

1. Introduction

Neuroeconomics is a very recent development within the research agenda called behavioral economics that has become rather popular at the end of the 20th century (Prelec, Camerer and Rabin, 2005; Glimcher, Camerer, Fehr and Poldrack, 2009). The departure point of this paper is the view that the appeal as well as the sources of criticism of neuroeconomics reside on its 'rhetoric and style'¹⁸ that

¹⁸ It is important to stress that we use the term rhetoric in a very broad fashion in order to highlight some bold claims neuroeconomists make to legitimate their

promise to revolutionize economic analysis by employing new brain imaging techniques that uncover heuristic processes and neural mechanisms underlying actual judgment and decision-making and even provide accurate direct measurement of economic models' preference and belief parameters.

Based on a methodological strategy that does not draw on prepackaged ideas of philosophy of science, we focus on the analysis of actual practice of neuroeconomics. This is because we aim to answer the following questions: (a) Why do a growing number of economists became interested in the intersections of economics, psychology and more recently neuroscience?; (b) How do neuroeconomists try to open up the black box of the human mind?; (c) What are the prospects in terms of predictive improvements and explanatory progress?; (d) What are the major challenges of neuroeconomics?, and (e) How do neuroeconomics contribute to the debate over individual rationality? In order to

practice. Given the modest purposes of the paper (that is to assess the possibilities and some challenges of neuroeconomics, we do not embark on a rhetorical analysis of neuroeconomics suggested by Deirdre McCloskey, Arjo Klamer and their fellows.

accomplish the foregoing tasks the remainder of the article is organized as follows.

Section 2 provides an interpretation of the emergence of economic analysis informed by the brain. Section 3 briefly presents some of the brain imagining techniques that contributed to the rapid development of neuroeconomics. Section 4 discusses some famous neuroeconomic experiments suggesting the neural foundations of social preferences and dynamically inconsistent choices. The main purpose is to have grounds for assessing the prospects of neuroeconomics in terms of predictive improvements and explanatory progress. Section 5 addresses some critical methodological issues related to the experimental strategies that guide neuroeconomic research. Section 6 analyzes the implications of neuroeconomics for the debate over rationality. Section 7 wraps the overall argument up and concludes that, despite some rhetorical exaggeration and methodological problems that characterize its practice, the young subfield of neuroeconomics is not a “hype” but genuine hope to contribute to the development of a theoretical approach with improved explanatory and predictive capabilities.

2. Neuroeconomics as ongoing efforts to open the black box of decision machinery

There is a long standing tradition suggesting that economics is a science that studies behavior as a relationship between ends and means (Robbins 1932, p. 16). In that perspective, economics as a study of the allocation problem need not envelop itself in the hazes of psychology and philosophy so as to unveil causal processes or mechanisms underlying actual judgments and decisions. A great deal of contemporary economics concentrates on developing a parsimonious and formally tractable framework (i.e. axiomatic revealed preference approach) that studies the abstract logic of choice (and takes the complex machinery of human values and decisions as a black box). The foregoing ideas are aligned with Milton Friedman's (1953) methodological view that the worth of any economic theory or model is to be judged by its predictive consequences (regardless of the descriptive accuracy or realisticness of its underlying assumptions).

Yet the growing literature of choice anomalies (i.e. recurrent and systematic evidence against the predictive power of the conventional utility framework) collected in the 1980s and

1990s inspired many objections to the standard economic theory of choice that partly explain the resurrection of psychological or behavioral economics. According to Tversky and Kahneman,

Deviations of actual behavior from the...model are too widespread to be ignored, too systematic to be dismissed as random error, and too fundamental to be accommodated by relaxing the... system.(1986, S272)

Contemporary behavioral economists are ready to add explicit psychological assumptions into their formal accounts if the reformed models can predict patterns of choice behavior already covered by the conventional approach as well as those empirical phenomena regarded as rational choice anomalies (Kahneman, Tversky, Knetsch and Thaler 1986). In this sense, the behavioral economist's aim to make accurate predictions is similar to the goal of neoclassical economics. Yet behavioral economists do not regard prediction as the only scientific purpose of economic models/theories. Their ongoing efforts to open up the black box of human preferences, values and decisions also reveal the goal of developing models or theories that provide an improved

understanding of how certain heuristic and mental processes bring about actual inferences and choices. Furthermore, behavioral economics takes individual choice experiments as important sources of inference about mental processes with major causal roles in behavior. Experiments are thought to help them to discern psychological explanations from conventional ones and to motivate the development of descriptively more accurate (i.e. realistic) models than those provided by neoclassical choice framework (Camerer and Loewenstein 2004, p. 7).

Neuroeconomics in turn can be interpreted as a development within behavioral economics (Camerer, Loewenstein and Prelec, 2005). It is driven by the conviction that the use of brain scanning techniques will contribute to further understanding the causal chains through which mental processes and neural entities and activities operate to bring about actual decision making behavior. According to neuroeconomists, recent advances in non-invasive brain imaging techniques open the door for the design of experiments in which economists and neuroscientists study the physical substrates of cognitive (and affective) processes underlying particular decision tasks and in so doing they

provide new data and methods to test and build alternatives to the (axiomatic) utility approach (Glimcher, Camerer, Fehr and Poldrack, 2009).

Putting it somewhat different, neuroeconomics provides a perspective to open the black box of the decision machinery by means of choice experiments that employ sophisticated brain imaging techniques. Such subfield purports to supplement behavioral economists' attempts to better explain and predict behavior. In a similar fashion, Colin Camerer claims:

Technology now allows us to open the black box of the mind and observe brain activity... these observations can only enhance the development of theories which are based on more accurate assumptions and make better predictions as a result (...) The neuroeconomic theory... replaces the perennially useful fiction of utility maximizing individual which has a single goal, with a more detailed account of how components of the individual – brain regions, cognitive control, affective system and neural circuits – interact and communicate to determine individual behavior (Camerer 2007, p. 28)

The next section shows how economists and psychologists draw on insights from the brain sciences so as to open up the black box and eventually better explain and predict behavior patterns regarded as anomalies and also those choices covered by standard theory.

3. The plurality of research methods characterizing neuroeconomics

Despite the fact that brain imaging techniques turn out to be the greatest tool of neuroeconomics, the field necessarily makes use of various research methods. This is because all of them have advantages and disadvantages (Camerer, Loewenstein and Prelec 2004, p. 557).

Most of the time brain imaging research consists of comparing people's performance under an experimental and a control task. The difference of neural activity detected under the foregoing situations serves as source of inference of what brain regions and neural activities physically execute mental processes underlying the foregoing tasks. Electroencephalogram (EEG), Positron Emission Tomography (PET) and functional resonance magnetic imaging (fMRI) are

the most popular research methods of neuroeconomics. There are also other procedures, such as hyperscanning, single neuron measurement, animal studies, electrical brain stimulation (EBS), transcranial magnetic stimulation (TMS) and so forth. In what follows, they are briefly presented and assessed.

3.1 Electroencephalogram (EEG)

This is one of the oldest techniques that rely on scalp electrodes so as to measure electrical activity from external areas of the brain. Despite the precise record of neural activity time, EEG carries the problem of poor spatial resolution and limitation to offer a satisfactory measure of interior brain processing activities.

3.2. Positron Emission Tomography (PET)

It is an improved technique that can measure neural activity in the inner areas of the brain and therefore provides a better spatial resolution and measure of brain activity than the EEG. PET measures blood flow, which seems to be a proxy of neural activity in particular areas of the brain. This research tool relies on a scanning procedure that require patients to

take an injection of a radioactive solution that emits positrons (i.e. positively charged electrons) in order to identify brain regions receiving more blood in response to glucose metabolizing). Yet this brain imaging development has to deal with the problem of poor temporal resolution.

3.3 Functional Resonance Magnetic Imaging (fMRI)

This is the most recent advance in technology for measuring brain processing activity. It does so by measuring oxygenated blood flow, which correlates with neural activities undertaken to deal with a particular task. The MRI is a way of recording neural activity with good spatial precision but it also faces the problem of weak temporal resolution. This limitation is often circumvented by the joint use of fMRI and EEG that can measure together inner and external signs of brain processing and therefore improved inferences about neural activity.

3.4 Hyperscanning techniques

This is an interesting advancement in the measurement of brain activity during interaction between people. It allows agents in MRI scanners in different locations to interact via the Internet and to have their brain activities simultaneously

measured. This opens room for the development of neuroeconomic experiments with various subjects resolving social tasks (Zak 2004)

3.5 Single neuron measurement

This is a procedure that requires a microelectrode in the subject's brain. Despite the capacity of good measurement of brain activity with spatial resolution, it is not often employed with humans, due to ethical issues. There are some exceptions to investigate animals and some surgical patients due to the brain damage risks that it carries.

3.6 Study of brain and behavior of individuals with neural damage

Some studies try to investigate the brain regions responsible for certain behaviors based on comparisons between choice behaviors of brain damage patients with normal subjects' performance (Bechara, Damasio, Tranel and Damasio 1997)

3.6 Transcranial Magnetic Stimulation (TMS)

This is a technique in which the use of magnetic field allows for temporary disruption or virtual lesion of certain areas of the brain under study. TMS promises to be a useful procedure to address the issue of mental causation. This is because one of the objections made to neuroeconomics is the difficulty in detecting the causal role of particular brain activations for decision-making (Glimcher, Camerer, Fehr, and Poldrack 2009).

Before assessing the potential for explanatory and predictive improvements that neuroeconomics promises to bring, it is important to stress that the foregoing research tools are regarded as complementary and try to overcome their relative weaknesses. The neuroscientist Paul Zak suggests that the plurality of methods reveal two main aims “(i) identifying the neural processes involved in decisions in which standard economic models predict behavioral well; and (ii) studies of anomalies where the standard model fail” (2004, p. 1740)

4. The Prospects of Predictive Improvements and Explanatory Progress

Experimental results from neuroeconomics inspire a debate over its predictive and explanatory implications. Given that the literature is becoming vast, we try to undertake the foregoing task by scrutinizing few neuroeconomic experiments that inform us about the neural foundations of prosocial preferences and neural structures physically associated with time conflicting preferences and intertemporally inconsistent choices. We concentrate on discussing Sanfey *et al.* (2003)'s study based on an ultimatum game and McClure *et al.* (2004)'s inquiry into time preference discounting. In addition, we claim that advances in neuroeconomics might be conducive to two notions of progress named unification and causal penetration (Mäki, 2004).

4.1 What neuroeconomics informs us about social preferences

There is significant experimental evidence suggesting that individuals are not only driven by self-regarding preferences and, under many contexts, they deviate from their material

self-interest on behalf of another's preferences and even react angrily at the perception of social exploitation or unfairness (Guth and Schmittberger 1982, Mansbridge 1990; Fehr and Gächter 2000). Neuroeconomists are willing to investigate the neural foundations of prosocial decision-making. Their goal is to uncover neural structures and brain processes or mechanisms that explain why people exhibit certain patterns of cooperative behavior that appear to be irrational or rational choice anomalies.

Inspired by behavioral economics, Sanfey and his collaborators designed a simple *ultimatum* game experiment (henceforth: UG) in which a player is asked to propose a money split of \$10 to another. If the responder accepts the offer, the sum is paid by the experimenter to each player. If she did not, both players end up with nothing.

The novelty of this UG experimental design is that 19 participants, who played the role of a responder, had their brains scanned through fMRI. The experimenters' aim was to study some neural and behavioral responses to fair and unfair money offers. Subjects inside a MRI scanner played the UG game with other players through a computer monitor and had

to play 30 randomly presented rounds (each lasting 36 seconds) – 10 with a computer partner, 10 with a human partner and 10 rounds that paid money by pressing a button. Experimental findings were very similar and gave extra boost to evidence collected by behavioral UG experiments. Each responder received five \$5 offers; two \$2 offers and two \$ 1 offers.

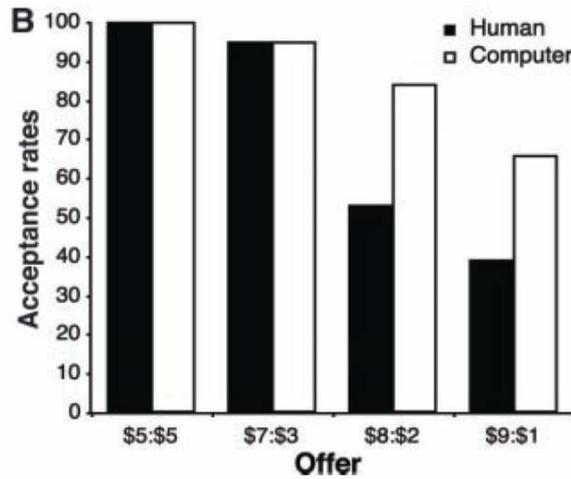


Figure 1 – Outcomes of the UG experiment.

Source: Sanfey, Rilling, Aronson, Nystrom and Cohen 2003, p. 1756

Sanfey et al found that individuals reject more unfair offers by human than those made by computer partners; what indicated stronger emotional responses to greedy human players than computers (Figure 1). They went on to test their hypothesis that unfair offers engage brain regions related to emotional and cognitive processing. According to them, the magnitude of activation in some neural structures probably explained differences in responder's decision to accept or reject positive money offers.

Based on a comparison between brain areas activated when agents were dealing with fair and unfair proposals, they detected that the magnitude of activation of bilateral anterior insula, dorsolateral prefrontal cortex (DLPC) and anterior cingulated cortex (ACC) was more intense for unfair proposals from human players. Neuroscience literature suggests that the bilateral anterior insula and the cingulated cortex are brain structures associated with emotion processing. Insula is related to emotional reactions of anger and disgust, whereas ACC is a conflict resolution area. DLPC, in turn, is a region responsible for physical execution of higher order cognitive processes and executive control, such as planning, goal achievement and evaluation. Activation of this area at the

perception of an unfair money offer might also be associated with representation and assessment of the goal of receiving money. Sanfey et al emphasized that activation of DLPC did not strongly correlate with acceptance rate and might not be crucial for prediction of money rejection. In addition, they found that rejected unfair offers are associated with greater activation of anterior insula than DLPC, whereas DLPC activation seems to bias behavior towards acceptance of positive money offers.

One interesting lesson drawn from the foregoing experiment is that anterior insula and DLPC seem to constitute the neural basis of emotional and higher order cognitive processes underlying decisions and choices of the UG type. This might support the neuroeconomist's prediction that people reject low offers when activation of insula is greater than the dorsolateral prefrontal cortex. In order to investigate the quality of prediction, Wout *et al.* (2005) and Knoch *et al.* (2006) employed the method of transcranial magnetic resonance to disrupt the DLPC and check its effect on people's acceptance of money offers in an ultimatum game. Their aim was to test the prediction that damaged DLPC turns off people readiness to reject unfair offers. The results confirmed the hypothesis.

Furthermore, they suggested that an assessment of the predictions made by neuroeconomic studies ought to “first establish parts of neural circuitry that implement a behavior; then stimulate or disrupt some parts and see if you can cause a behavioral change” (Camerer 2007, p. C37)

Other experiments also detect neural foundations of social preferences and cooperation. For instance, McCabe *et al.* (2001) suggest that individuals who cooperate more have increased activation in a brain region called Brodmann Area 10, often regarded as the mind-reading circuitry. In a UG study comparing the performance of normal and autistic participants (i.e. individuals that are thought to have impaired Brodmann area 10), Hill and Sally (2002) found that autists, playing the role of a proposer, offer small (if any) sums of money due to their difficulty in imagining others’ emotional reaction to an unfair money split proposal. Based on some neuroeconomic studies, Paul Zak and his collaborators find that a hormone called oxytocin plays an important role in trust behavior (Zak *et al.* 2003, Zak *et al.* 2005). In order to test the conjecture, Kosfeld and others (2005) designed an experiment in which half of the participants playing a trust game received an oxytocin dose and another half a placebo.

Each player could choose whether to keep money or to invest in a trust fund. If he did the latter, the sum of invested money would double. In this case, the trustee (the responder subject) would have to choose how much to keep and to pay. Empirical results give support to the experimenters' prediction that oxytocin would increase trust behavior.

4.2 Neuroeconomics' contributions to the understanding of choice over time

Neuroeconomics challenges some theoretical concepts and assumptions underlying standard economic models of intertemporal choice. In the conventional approach, the way in which individuals undertake their utility tradeoffs is represented in terms of the concept of discount rate. The latter amounts to the rate at which future utility are discounted based on the period when consumption of particular goods happens. Behavioral economists and psychologists have collected various pieces of evidence suggesting that the assumption that individuals discount all future utilities at a constant rate constrains the predictive and explanatory powers of the conventional economic model of intertemporal choice (Loewenstein and Elster 1992, Frederick et al 2003,

Read and Baumeister 2003). Alternatively, they propose a variant of the utility model in which utilities are discounted in a non-constant fashion. One of the starting points of neuroeconomists interested in the topic of choice over time is the idea that an improved understanding of how the brain works will contribute to better accounts of actual choices. Neuroeconomists attempt to unpack the concept of time preference and discounting and provide better predictions of when people exhibit dynamically inconsistent behaviors and explanations of why and how this happens. In Camerer, Loewenstein and Prelec (2005)'s views, "neuroscience points to some deficiencies in the way economists currently model intertemporal choice and also suggests directions for future modeling' (p. 43)

Neuroeconomic experiments seem to give extra credence to the behavioral idea that intertemporal decision-making consists of two mental systems, a cognitive (deliberate, controlled) and an affective (impulsive, automatic). In an interesting experiment in which subjects choosing among various monetary gains spread over distinct periods of time had their brains scanned through MRI, Samuel McClure and colleagues investigate the conjecture that intertemporally

inconsistent behavior is due to the competition between the two systems respectively correlated with prefrontal cortical regions and the limbic system. The experimenters try to investigate whether and why individuals ascribe strong importance to immediate consumption at the present but prefer to delay gratification as time goes by. More specifically, they aim to find out whether this phenomenon **has a neural foundation and can be represented by a quasi-hyperbolic discounting function.**¹⁹ The latter suggests that the discounted value of a consumption, u , at delay $t=0$ is u and $\beta \delta^t \cdot u$ for $t > 0$, $0 \leq \beta \leq 1$ and $\delta \leq 1$.²⁰ McClure and his collaborators hypothesize that parameters β and δ represent behavioral predispositions activated by brain regions responsible for impulsiveness and areas responsible for careful thinking.

¹⁹ Quasi-hyperbolic discounting suggests that an individual has very high discount rates for present vis-à-vis future rewards and lower rates for choices that are time-dated in the future. For details, see Laibson (1994).

²⁰ The inverse of δ is the value people give to immediate reward, while β values a given delay regardless of time.

Empirical results show that the subcortical area of the limbic system was strongly activated by alternatives of immediate (lower) rewards. Regions of the prefrontal cortex and parietal cortex, in turn, are activated in a uniform fashion regardless of the time delay of monetary rewards. Based on a comparison between relative activity of the affective and higher order cognitive systems, the experimenters found evidence congruent with behavioral economist's prediction. Choice patterns that involve immediate reward (no delay) is associated with higher activation of limbic neural structures - ventral striatum, medial frontal cortex and cingulate cortex (i.e. physical counterpart of the \square system). Choices that reveal individual preference for delayed gratification are correlated with more activation in lateral orbitofrontal cortex, dorsolateral cortex, and parietal cortex often regarded as structures physically responsible for higher order thinking and deliberation (\square system).

One implication of neuroeconomic studies of intertemporal choice is that it offers an improved account of why individuals often deviate from predictions made by the constant discounting model. Choice over time in the real world involves two competing neural systems (cognitive and affective) that

might explain and predict dynamically inconsistent behavior often considered intertemporal choice anomalies.

4.3 Neuroeconomics and explanatory progress

The foregoing discussion motivates us to address an important philosophical question concerning the contributions of neuroeconomics to the pursuit of explanatory progress. In a sense, neuroeconomics might be progressive to the extent that it relies on new research methods of brain imaging so as to address more empirical questions that standard economic analysis can (Bernheim 2009). In Paul Zak's words,

(...) Neuroeconomic research will lead to the building of models that predict economic and social behaviors better and that are grounded in neurobiology. This will allow economists to answer fundamental questions they are unable to address now such as: why do two individuals faced with the same information and incentives make different choices? Why does the same individual sometimes make choices that are inconsistent? ((2004, p. 738)

Inspired by Mäki (2004), we claim that the reformist methodological and theorizing strategies of neuroeconomics

to open up the black box of the decision machinery reveal researchers' pursuit of explanatory progress interpreted as scope unification or unification and also in terms of causal articulation or penetration.

The first notion of progress refers to the potential for a given set of explanatory elements or schemes to accommodate a growing number of *explanandum* phenomena. Putting somewhat different, it amounts to a model or theory's capacity to expand its scope or set of explained phenomena. In this case, a model or theory that is progressive is able to "explain much by little". Neuroeconomists investigate how certain neural processes and mechanisms give rise to choice behavior with the aim to develop a general framework with unifying properties, i.e. able to provide an explanatory scheme under which apparently different and unrelated phenomena can be subsumed. As Camerer, Loewenstein and Prelec put it:

A third payoff from neuroscience is to suggest that economic choices which are considered different in theory are using the same brain circuitry. For example, studies... found that insula cortex is active when players in ultimatum games receive low

offers, when people choose ambiguous gambles or money, when people see faces of others who have cooperated with them, and in players who are poor strategic thinkers. This suggests a possible link between these types of games and choices which would never have been suggested by current theory” (2004, p.574)

It is important to stress that a unified approach might bring some loss in simplicity because its capacity to accommodate an increased number of *explanandum* phenomena results from an increase in the set of explanatory items and variables. Yet the potential for explanatory gains is thought to outweigh its costs in terms of simplicity (Muramatsu 2006, p. 85)

Progress as a matter of causal articulation refers, in turn, to the potential for developing a theory or model that succeeds in selecting the explanatory items playing major roles in the causal production of the chosen *explanandum* phenomenon. This second version of progress is concerned with specifying how exactly items of the explanatory set interact to bring about a certain phenomena under study and eventually attempts to locate the place of the explanandum and its major constituents in a complex world of causal dependencies

(Muramatsu 2006, pp. 84-85). Although a great deal of neuroeconomic research provide more solid evidence for correlation than for causation, neuroeconomists and behavioral economists explicitly reveal their concern with causal articulation of their revised models of choice. To Camerer and his colleagues,

A forth potential payoff from neuroscience is to add precision to functions and parameters in standard economic models (...) The point is that knowing which neural mechanisms are involved tell us something about the nature of the behavior. For example, if the oxytocin hormone is released when you are trusted, and being trusted sparks reciprocation, then raising oxytocin exogenously could increase trustworthy behavior. (...) The basic point is that understanding the effects of biological and emotional processes like hormone release and mood will lead to new types of predictions about how variations in these processes affect economic behavior (Camerer, Loewenstein and Prelec 2004, p. 575)

Although progress as unification can come with or without causal articulation, neuroeconomists emphasize that their

goal is to pursue both types of progress. Paul Glimcher and Aldo Rustichini rhetorically frame the issue:

Economics, psychology and neuroscience are converging today into a single, unified discipline with the ultimate aim of providing a single, general theory of behavior. Economists and psychologists are providing rich conceptual tools for understanding and modeling behavior, while neurobiologists provide tools for the study of mechanism. The goal of this discipline is thus to understand the processes that connect sensation and action by revealing the neurobiological mechanism by which decisions are made (p.447) (...) The goal of the emerging neuroeconomic program will have to be a mechanistic and behavioral explanation that transcends the explanations available to neuroscientists, psychologists and economists working alone” (Glimcher and Rustichini, 2004, p. 452)

As we can see, the practice of neuroeconomics is driven by the view that improved explanations of complex phenomena like decision-making, strategic thinking and social exchange are dependent on a clearer grasp of how affective, cognitive processes and neural mechanisms work together to produce

behavior patterns (anomalies included). Despite the interesting available experimental results from neuroeconomic research, it is important to understand some objections that have been recently made to it.

5. Doubts and challenges posed to the practice of neuroeconomics

Not only does the infant agenda of neuroeconomics have big advocates, but encounters important critics. We claim that this is largely due to some neuroeconomists' 'rhetoric and style' that promise to revolutionize and unify economics, psychology and neuroscience (Rubinstein 2008) and even to provide perfect accurate measurement of models' preference and belief parameters. In our perspective, most of the objections to this new subfield relate to methodological issues that this section tries to address.

5.1 Neuroeconomics provide completely new theoretical constructs and methods that revolutionize economics

In their rather eloquent article entitled "Neuroeconomics: how neuroscience can inform economics", Camerer *et al.* argue that their subfield will provide completely new theoretical

constructs, direct measurements of preferences, thoughts and feelings and a completely different understanding of economic decision-making behavior (p.10). Yet a growing number of economists show some skepticism and raise doubts about the actual success of current neuroeconomics in accomplishing all that (Gul and Pesendorfer 2008, Harrison 2008, Bernheim 2009).

The first criticism points out to the very domain of positive economics and the relevance of brain imaging for its development. In Gul and Pesendorfer (2008)'s view,

Neuroscience evidence cannot refute economic models because the latter makes no assumptions about physiology of the brain (p.22) ... Economic phenomena consists of individual choices and their aggregates and do not include hedonic values (...) Brain imaging data are of a radically different form than are typical of economic data. If the prediction of great advances in brain science turns out to be correct, they will certainly be accompanied by theoretical advances that address the particular data in that field. It is unreasonable to require those theories to be successful at addressing economic data. By the same token, the

requirement that economics simultaneously account for economic data and brain imaging data places an unreasonable burden on economic theories (p.33)

Even though it is fair to say that economic data and brain imaging data can refer to different phenomena, we doubt that one can say *a priori* that neuroscientific research is non-relevant to economics. There is little consensus about whether the revealed preference approach is enough to meet the ambitious scientific goals of a discipline like economics that has a highly expanding scope. Instead, we opt for suggesting that exploration of new kinds of data and research methods might payoff especially because standard data and methods seem to be insufficient to measure parameters, to relate relevant behavioral variables, to better specify and to select new or available models.

A related criticism has to do with the exaggerated 'rhetoric' of neuroeconomics. There are doubts that so far neuroeconomics offer completely new theoretical constructs and direct measurements of preference parameters. Many concepts and experimental techniques clearly descend from the literature of judgment and decision-making and

psychological economics (Harrison 2008). For instance, neuroeconomists rely on available notions of state-dependent preferences, quasi-hyperbolic time discounting and social preferences. They also rely on laboratory individual choice experiments to test their hypotheses. The distinguishing feature is that neuroeconomic studies reform behavioral economic experiments by incorporating recent brain imaging methods. This leads us to other critical remarks about the epistemology of neuroeconomic experiments.

5.2 Challenges posed to experimental practices

Just like experimental and behavioral economics, the subfield of neuroeconomics is sensitive to criticisms associated with the very design of experimental settings, their incentive structures and quality of controlled variables. Neuroeconomists also need to deal with two recurrent problems dubbed external and internal validity. The former refers to the difficulty in generalizing results from the laboratory (e.g. scanned brains) to the outside world. Internal validity amounts to the problem of knowing whether a particular activity, process or mechanism (activation in

prefrontal cortex and limbic system) detected in the experiment is correct.²¹

External validity is a non-negligible problem provided that most neuroeconomic experiments involve a very small number of participants (Harrison 2008). Some overly enthusiastic neuroeconomists fail to acknowledge that this is a shortcoming and make efforts to investigate the effect of the size of their sample on results. For instance, Bhatt and Camerer (2005) simply claim that addition of more participants in their experiments would not change the outcomes. Some neuroeconomists try hard to weaken the objection of external validity by stressing that the brain is not the unit of analysis, but patterns of neural activations during some decision tasks (Glimcher, Camerer, Fehr, and Poldrack 2009).

Internal validity is a no less problematic issue than external validity. It poses one important challenge to the very development of neuroeconomics: do our standard

²¹ For details, see Guala (2005)

econometric toolbox deal well with brain imaging data? Harrison is rather skeptical. He warns us:

Brain scans do not light up like Christmas tree without lots of modeling assumptions being made. Recognizing this fact is not meant as a way of invalidating ...or research goals of the exercise, but an attempt to be sure that we understand the extremely limited extent to which modeling and estimation errors are correctly propagated throughout the chain of inferences (...) The overall implication is that one should be concerned that there is a significant understatement of standard errors on estimates and effects, implying a significant overstatement of statistical significant differential activation (2008, p. 13)

A careful look at some experiments lead us to suggest that they identify some interesting statistical associations (correlations) between variables (parameters) but there is no room for an indisputable move to a 'causation talk'. Take as examples research on neural foundations of quasi-hyperbolic discounting and prosocial preferences (discussed in the previous section).

McClure and colleagues investigate whether individual preferences for immediate lower rewards are related to automatic affective processing areas of the brain, while choices over delayed (higher) gratification are related to controlled cognitive processing regions. The experimenters do not suggest that the validity of the quasi-hyperbolic discount function ($\beta \delta^t u$) depends on activation of two different brain areas. They simply suggest that a differential pattern of brain activity would be consistent with the quasi-hyperbolic specification. In this case, we cannot draw the conclusion that quasi-hyperbolic discounting is a physiologically true representation of individual structure of intertemporal utility tradeoffs because limbic neural structures (ventral striatum, medial frontal cortex and cingulate cortex) constitute the physical substrate for the β system, whereas activation in lateral orbitofrontal cortex, dorsolateral cortex, and parietal cortex (brain structure often associated with higher order thinking and deliberation) cause individual preference for delayed high gratification and compose the brain foundation of the δ system.

Experiments that investigate the neural basis of trust behavior also seem to face the problem of internal validity. Results

suggesting the influence of oxytocin on trust behavior do not allow for inferences about what caused first move players to send more money to a social fund and second move players to send more money back. There could be other driving forces behind this instance of prosocial choices, such as altruism, perception towards risk, understanding of the experimental design, etc.

The abovementioned problems highlight various methodological issues that neuroeconomics ought to address and overcome so as to succeed in developing a genuine explanatory scheme that improves our understanding of decision-making in the real economic world. Other subfields like experimental economics have dealt with similar difficulties and go on working hard to find their own ways to triumph over them.

6. Implications for the debate over rationality

Neuroeconomic research draws on insights from evolutionary biology so as to hypothesize that brains are equipped with domain specific information-processing systems (for instance, the affective and the cognitive) that evolved to deal with

particular tasks. Neuroscientists' empirical results indicate that effective operation of the decision machinery requires controlled deliberative neural mechanisms as well as automatic ones.²²

More importantly, they converge on psychologists' and behavioral economists' view of bounded rationality (henceforth: BR).

It is important to stress that the broad notion of BR is used to criticize conventional economic explanation of rational choice as (expected) utility maximization (Simon 1955, 1956, 1997; Chorvat and McCabe 2005). Most economists regard BR as 'optimization under constraints'; they draw on the psychologist's idea that actual individuals have limited computational capabilities and therefore make decisions based on simple choice heuristics or routines (e.g. satisficing) that economize on cognition. Nonetheless, a careful reading of Simon's work lead us to oppose to the foregoing vision of

²² For instance, individuals who have impaired neural capacity for emotion processing of information about risky prospects make worse decisions than normal people (Bechara, Damasio, Damasio and Lee 1999).

BR. Optimization under constraints seems to require even more cognitively demanding information processing tasks than those idealized by the conventional economic model of rational choice (Muramatsu and Hanoch 2005, p. 210). In response to that, we prefer an alternative account of bounded rationality that is inspired by Simon's understanding of choice behavior as a result from the operation of a mechanism constituted by a pair of scissors, whose blades are the agent's computational facilities and the structure of the task environment (Gigerenzer and Selten 2001). This alternative interpretation of BR in terms of an (adaptive) toolbox of heuristics processes or neural mechanisms hypothesizes that the thinking brain has evolved to resolve domain-specific tasks.

The notion of ecological rationality implies that performance is evaluated in terms of the match between a heuristic and the structure of the environment. In other words, a heuristic will be ecologically rational to the degree that it is adapted to the informational structure of a specific environment and gives rise to a satisfactory course of action (Muramatsu and Hanoch, p.211).

6.1 Neuroeconomics gives boost to the notion of ecological rationality

In our understanding, neuroeconomics entertain the vision of BR as ecological rationality to the extent that it is built on the hypothesis that the brain is endowed with two domain specific information processing systems - controlled (cognitive) and automatic (affective) – that work together to produce of judgment and decision-making behavior. In such perspective, experimental findings also suggest that individuals are boundedly rational agents that often exhibit patterns of behavior deviating from predictions by standard economic theory (based on the utility optimization criterion).

It seems that one of the great contributions of brain imaging studies to contemporary economic analysis is to provide other types of data and methods that help economists to better explain why and under what circumstances individual choices are consistent with the standard model of rational choice and why and under what environments anomalies occur. According to the notion of ecological rationality, individual choice might exhibit patterns of choice behavior deviating from standard theory's prediction because some activated

neural mechanisms or heuristic processes (adapted to a previous task context in which it evolved) do not perform well in the contemporary task environment. This need not result from an individual's lack of a particular neural mechanisms or heuristics that bring about an optimal (rational) outcome. Ecologically irrational behavior can happen because some specific neural mechanism that execute agent's mental capability for opting for the best choice alternative remains dormant or is not strong enough to neutralize activation of competing mechanisms (Vromen 2007).

With this all mind, we are inclined to suggest that the conjecture that our information-processing brain is equipped with specific neural systems that sometimes supplement and other times compete with each other offers a way of explaining (understanding) why and when rational as well as suboptimal choice behaviors emerge. More fundamentally, the notion of ecological rationality guiding the practice of neuroeconomics sheds extra light on the limited explanatory capabilities of standard accounts of constructivist rationality (i.e. explanations deduced from a compact set of utility axioms

and probability rules), and on the possibility of regarding emotions as processes with major roles in production of some rational economic decisions (Muramatsu 2006).

7. Final Remarks

This article draws on some insights from economic methodology so as to scrutinize the possibilities and limitations of neuroeconomics. Based on a thorough analysis on the motives behind economists' willingness to open up the decision machinery and the potential implications of using brain imaging data for testing concepts and building alternative models of choice, we are led to conclude that neuroeconomics is still in its infancy but can contribute a lot to the behavioral economist's goal of developing a general theoretical scheme that better predict and explain the complex phenomenon of decision-making behavior and its anomalies.

Despite some current rhetorical exaggeration and debatable experimental practices, we argue that neuroeconomics is not a mere 'hype' but genuine hope to achieve predictive and explanatory progress. Yet the validity of our claim strongly depends on whether near future neuroeconomists recognize

explicitly some of their methodological difficulties and work hard to make precise theoretical and empirical claims about the (causal) connections between neural structures or mechanisms and behavioral outcomes. Furthermore, neuroeconomists are in need of specifying their moderate physicalist view that mental phenomena are at bottom physical brain phenomena and what consequences it brings to the structure of economic explanation and discourse.

BIBLIOGRAPHY

Bechara, A., Damasio, H., Damasio A., Lee, G. (1999) 'Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making,' *Journal of Neuroscience*, 19, pp. 5473-5481

Bechara, A., Damasio, H., Tranel, D. and Damasio, A. (1997) 'Deciding Advantageously Before Knowing the Advantageous Strategy,' *Science*, 275, pp. 1293-1294

Bernheim, D. (2009) 'The Psychology and Neurobiology of Judgment and Decision-Making: what's in it for Economists?,'

In: Glimcher, P., Camerer, C., Fehr, E. and Poldrack, R.A. (eds.), *Neuroeconomics: decision making and the brain*. New York: Academic Press

Bhatt, M. and Camerer, C. (2005) 'Self-Referential Thinking and Equilibrium as States of Mind in Games: fmri evidence,' *Games and Economic Behavior*, vol. 52 (2), Aug., pp. 424-459

Camerer, C. (2007) 'Neuroeconomics: using neuroscience to make predictions,' *The Economic Journal*, 117 (March), pp. C26-C42

Camerer, C., Loewenstein, G. and Prelec, D. (2005) 'Neuroeconomics: how neuroscience can inform economics,' *Journal of Economic Literature*, vol. XLIII, March 2005, pp. 9-64

Chorvat, T. and McCabe, K. (2005) 'Neuroeconomics and Rationality,' *Chicago Kent Law Review*, 80, pp. 101-121

Fehr, E. and Gächter, S. (2000) 'Cooperation and Punishment in Public Good Experiments,' *American Economic Review*, 90, pp. 980-994

Frederick, S., Loewenstein, G., O'Donoghue, T. (2003) 'Time Discounting and Time Preference: a critical review'. In: Read, D. and R. Baumeister (2003) *Time and Decision: economic and psychological perspectives on intertemporal choice*. New York: Russel Sage Foundation

Gigerenzer, G. and Selten, R. (2001) *Bounded Rationality: the adaptive toolbox*. Cambridge: MIT Press

Glimcher, P. and Rustichini, A. (2004), 'Neuroeconomics: the consilience of brain and decision,' *Science*, vol. 306, 15 October 2004, pp. 447-452

Glimcher, P., Camerer, C., Fehr, E. and Poldrack, R.A. (2009), *Neuroeconomics: decision making and the brain*. New York: Academic Press

Guala, F. (2005) *The Methodology of Experimental Economics*. New York: Cambridge University Press

Gul, F. and Pesendorfer, W. (2008) 'The Case for Mindless Economics,' In Caplin and Schotter (eds.), *The Foundations of Positive and Normative Economics: a handbook*, New York: Oxford University Press

Guth, W., Schmittberger, R. and Schwarze, B. (1982) 'An Experimental Analysis of Ultimatum Bargaining,' *Journal of Economic Behavior and Organization*, 3, pp.367-388

Harrison, G. W. (2008) "Neuroeconomics: a critical reconsideration," Working Paper 08-01 Department of Economics, University of Central Florida. Published by Economics and Philosophy, 24, 2008, [Online] available from [www.bus.ucf.edu/wp/working %20papers/2008/08-01%20Harrison.pdf](http://www.bus.ucf.edu/wp/working%20papers/2008/08-01%20Harrison.pdf) ; accessed on August 15, 2008; Internet

Hill, E. and Sally, D. (2002) 'Dilemmas and Bargains: Autism, Theory of Mind, Cooperation and Fairness,' *University College London Working Paper*, [Online] available from <http://ssrn.com/abstract=407040>; Internet, accessed on September 15, 2008

Knouch, D., Pascal-Leone, A., Meyer, K., Treyer, V. and Fehr, E. (2006) 'Diminishing reciprocal fairness by disrupting the right prefrontal cortex,' *Science Express*, vol. 314 (5800), pp. 829-832

Kosfeld, M. Heinrichs, M., Zak, P., Fischbacher, U. And Fehr, E. (2005)'Oxytocin increases trust in humans,' *Nature*, vol. 435 (7045), 2nd of June, pp. 673-676

Laibson, D. (1994) *Essays in Hyperbolic Discounting*, PhD dissertation, Massachusetts Institute of Technology (MIT)

Loewenstein, G. and Elster, J. (1992) *Choice over Time*. New York: Russel Sage Foundation

Mäki, U. (2004) 'Theoretical Isolation and Explanatory Progress: transaction cost economics and the dynamics of dispute,' *Cambridge Journal of Economics*, 28, pp. 319-346

Mansbridge, J. (1990) *Beyond Self Interest*. Chicago: University of Chicago Press

McCabe, K., Houser, D., Ryan, L., Smith, V. and Trouard, T. (2001)'A functional imaging study of cooperation in two person reciprocal exchange' *Proceedings of the National Academy of Sciences, USA*, 98, 11832-11 35

McClure, S. M., Laibson, D. I., Loewenstein, G., and Cohen, J. (2004) 'Separate Neural Systems Value Immediate and

Delayed Monetary Rewards,' *Science*, vol. 306, 15 October, pp. 503-507

Muramatsu, R. (2006) *Emotions in Action: an inquiry into the explanation of decision making in the real economic world*. PhD dissertation, Erasmus Institute for Philosophy and Economics (EIPE), Erasmus University of Rotterdam

Muramatsu, R. and Hanoch, Y. (2005) 'Emotions as Mechanism for Bounded Rational Agents: the fast and frugal way,' *Journal of Economic Psychology*, 26 (2005), pp. 201-221

Read, D. and R. Baumeister (2003) *Time and Decision: economic and psychological perspectives on intertemporal choice*. New York: Russel Sage Foundation

Rubinstein, A. (2008) 'Comments on Neuroeconomics'. Also published by *Economics and Philosophy*, 24 [Online] available from <http://www.arielrubinstein.tau.ac.il/papers/neuro.pdf>; accessed on August 15, 2008; Internet

Sanfey, A. G., Rilling, J. K., Aronson, A. L., Nystrom L.E. and Cohen, J.D. (2003) 'The Neural Basis of Economic Decision-

Making in the Ultimatum Game,' *Science*, vol. 300, 13, June, pp. 1755-1758

Simon, H. (1955) "A behavioral model of rational choice," *Quarterly Journal of Economics*, 69, pp. 99-118

Simon, H. (1956) "Rational Choice and the Structure of Environments," *Psychological Review*, 63, pp. 129-138

Simon, H.(1997) *Models of Bounded Rationality: empirically grounded economic reason*, vol. III. Cambridge: MIT Press

Vromen, J.J. (2007) 'Neuroeconomics as a Natural Extension of Bioeconomics: The Shifting Scope of Standard Economic Theory,' *Journal of Bioeconomics*, 9, pp. 145-167

Wilson, E. (1998) *Consilience*, New York: Knopf

Zak, P. J. (2004) 'Neuroeconomics', *Phil. Trans. R. Soc. Lond. B*, 359, pp. 1737-1748

Zak, P. J., Kurzban, R. and Matzner, W.T. (2005) 'Oxytocin is associated with human trustworthiness,' *Hormones and Behavior*, vol. 48 (5), December, pp. 552-557

**NEUROECONOMICS:
ISSUES OF SCIENTIFICITY AND
INTERDISCIPLINARITY**

Uskali Mäki
Academy of Finland

I'm going to supplement Roberta's nice outline by making a few remarks about a selected set of further issues. These issues have to do with the scientific status of neuroeconomics and its interdisciplinary character. Those are issues that are indeed on the table now. There are diverging opinions on what makes good science and on whether economics might benefit – so as to improve its scientific performance - from importing ideas from other disciplines.

Neuroeconomics is a new research field, it has been around for ten years or so, and it is already under severe attack. This is a very interesting phenomenon. Such a response did not happen at the same scale with experimental economics and many other new streams within economics; they were not met with such a severe attack from early on, by more conventional economists. Of course, the rise of neuroeconomics has been

very fast. Just look at the number of practitioners of neuroeconomics and the number of research centers that have been established in the recent years. The number of publications is booming. Perhaps this speed has played a role in triggering critical reactions. Yet it seems obvious that no counter-attack can stop it.

Just recently, there have been two publications that document these early debates around neuroeconomics. One is a special issue of the journal *Economics and Philosophy* (October 2008), and the other is an edited volume (with an odd title) *The Foundations of Positive and Normative Economics. A Handbook* (Oxford University Press 2008). These two collections of essays give a good idea about some of the key issues and about the prevailing arguments and attitudes.

This is a very interesting situation. It is a great opportunity for philosophers of economics to become active and make a difference as this is one of those historical occasions on which practicing economists themselves are sufficiently inspired to engage in this sort of philosophical reflection on foundational issues in their own discipline. These situations arise every now and then, but they surely are not a permanent or regular

characteristic of actual scientific practice. Whether it does arise depends on further characteristics of the context. There must be some sufficiently serious challenges that cannot be easily ignored or escaped and that require rethinking some foundational issues. Obviously, this is now happening. Many practicing economists seem to feel like being forced to take a stance on some deep issues that do not usually arise in their daily professional activities. They now explicitly formulate ideas about the identity and boundaries of economics and about the scientific values and standards and goals that are supposed to drive and control research and controversy in their discipline. This is a golden opportunity for those specializing in philosophy of economics to contribute to these debates.

Critics often blame economists for not paying attention to whatever happens in neighbouring disciplines, that economics is too insulated and self-contained, that it should rather be more engaged in interdisciplinary interactions. The official self-image of economics is that economics is an autonomous discipline, economists don't need to know what happens elsewhere in other disciplines. But, of course, this is just the official self-image, while in practice economics has imported

many of its ideas across disciplinary boundaries - just think of ideas coming from mathematics and statistics and physics. Without these imported ideas, economics wouldn't be what it is now. But recently, economics has become more active not only in importing but also in exporting ideas across disciplinary boundaries, so there has been action in both directions, giving and taking.

When thinking of interdisciplinary relations and transfer of ideas, it is useful to make a distinction between horizontal and vertical directions. The vertical one is between so-called levels. Ontologically speaking, there are supposed to be levels such as those of atoms, molecules, brain cells, individual minds, and social institutions. In science we have physics, chemistry, biology, psychology and then the social sciences, roughly corresponding to these levels and investigating them.

As to the horizontal relations, here the interactions (or the lack thereof) between disciplines take place in addressing a given level, say the social level. Social sciences highlight social phenomena in various ways, and they sometimes borrow ideas from one another. The dominant horizontal trend is that

of economics exporting ideas to other social sciences, political science, law, geography, sociology, etc. This is the famous economics imperialism, the use of notions such as rational choice in the market for explaining social phenomena that traditionally have been conceived as non-economic (see Mäki 2009). Economics is spreading its influence across other social sciences at roughly the same horizontal level – or a couple of neighbouring levels - as it were.

The issue about neuroeconomics is concerned with vertical relations between levels. Neuroeconomists themselves often say they are in the business of opening the black box of the individual rational agent (or of the human mind or the human brain). This is a matter of moving between levels. By looking into the brain they claim to move from the mental to the physiological – to a “deeper” level in the vertical direction.

It may be granted that much scientific progress occurs by way of opening what was a black box and by identifying and describing some mechanism inside. This is also how many neuroeconomists see their own contribution. They believe this will provide us with benefits such as improved predictions or deeper understanding or better confirmation of theories. Then

there are others who say that there is no need to open this particular black box: its opening will not yield any useful information given the alleged tasks of economics.

In case we agree that this box had better be opened, the next question is to ask how its contents should be illuminated. What kind of theoretical resources, drawn from which discipline, should be exploited in illuminating the darkness? One would expect that an obvious answer is: the resources of neurobiology. It is true that the *techniques* of brain research have been employed, but this field has turned out to be relatively less resourceful insofar as *theory* is concerned. This observation has encouraged some neurobiologists to use economic theory as such a resource for illuminating the box of the brain in terms such as the market for neural activity. Regardless of whether one wants to view this as yet another manifestation of economics imperialism (now in a vertical direction), it does manifest a split into two branches of neuroeconomics. One branch begins from economics and the wish to improve its performance by looking into the human brain, consulting neurobiology for assistance. The other branch begins from neurobiology, wishing to sharpen its

theoretical structure, and seeks to do this by consulting economics for theoretical guidance.

Let us then talk about the issue of scientificity, the status and quality of a piece of inquiry as science. Is it science or non-science, and if it is science, is it good or bad science? There are conflicting perceptions about neuroeconomics in regard to these questions. Some may say that finally economics is becoming a hard science. It used to be a soft science but now we start measuring things much as the natural sciences do and thereby making economics a hard science. By contrast, others say that neuroeconomics is sloppy and lacks rigour in the use of concepts, in collecting, sharing and analyzing data, and in drawing conclusions – in a word, it violates the standards of good science.

Some of these issues about neuroeconomics can be understood from the point of view of familiar controversies around interdisciplinarity. Disciplinarity amounts to conforming to the rules and conventions of an established scientific discipline, much as Kuhnian normal science is disciplined by the dicta of a paradigm. Interdisciplinarity is a matter of crossing disciplinary boundaries and thereby possibly moving

beyond the tight grip of the dicta of any particular discipline. The images of both disciplinarity and interdisciplinarity are mixed and contested. From one perspective, disciplinarity is something not so praiseworthy: it is static, conservative, elitist, inflexible, narrow, boring, averse to innovation -- whereas interdisciplinarity is good, because it is dynamic, open, flexible, broad, democratic, exciting, innovative, etc. But then, looking at these things from another point of view, the images are turned upside down: disciplinarity appears as recommendable since it comes with qualities such as rigorous, competent, systematic, professional, orderly, and stringent standards -- whereas interdisciplinarity is associated with attributes such as vague, sloppy, dilettante, unorganized, messy, floating standards, and so on. The two perspectives provide very different images of interdisciplinarity. These are also the two major images of neuroeconomics.

The rhetoric of inquiry is also part of the picture. With respect to many audiences, neuroeconomics has been persuasive, enhanced by successful self-promotion. It has done very well in attracting excited attention as well as in fund raising compared to many other fields in economics. Unsurprisingly, the critics have often blamed neuroeconomics for rhetorical

excesses, for salesmanship promising and pretending to deliver more than it can.

Are the complaints about the questionable performance of neuroeconomics justified? One way of answering this is to appeal to the interdisciplinary character of neuroeconomics. If your work proceeds strictly within the confines of a discipline, if you stick to the established rules of your discipline, you are judged to be competent and able to observe what are perceived to be high standards of quality. But if you go interdisciplinary and start stepping outside the boundaries of your own discipline, the standards of quality may start moving too. Your work may conform to the rules and standards of a neighbouring discipline or of no existing discipline at all. If we read the notion standard in a certain way, it is typically not the case that any two disciplines would share exactly the same specific standards (even if it happened to be the case that all of science were to share the same abstract principles). Standards and their interpretations are typically discipline-specific, and therefore certain forms of interdisciplinarity may require violating some standards of a particular field. For this reason we may excuse some problems in neuroeconomics at least on a temporary basis. There may come a time in the

future such that neuroeconomics will have established itself as a field or discipline or has acquired the characteristics of a discipline such that it will have adopted disciplining standards to regulate its own practices. And then, these problems may vanish.

Another related way to answer the question about whether the complaints are justified is to refer to novelty. Neuroeconomics is a new field of inquiry, an immature freshman line of research taking its early steps. It obviously has some way to go to become a mature field. And it seems to be moving in that direction rather than making regress. While looking for new results and new information about the brain, it is simultaneously also searching for standards of quality that can be used for assessing its own performance. These developments go together. Substantial progress and progress regarding standards go hand in hand, so I think this provides some grounds for temporarily excusing or bracketing some of the problems of neuroeconomics' performance. (Naturally, this should not go too far so as to undermine the obvious principle that only by paying attention to flaws can those flaws be corrected!) In regard to accusations of rhetorical excesses, one possible response is to appeal to novelty. In order to

survive and to have a chance of making progress, a new and entrepreneurial research endeavour must attract attention, generate hopeful expectations, and raise resources. It is the task of excited rhetoric to secure these preconditions.

Ariel Rubinstein is also worried about low standards, but he explains this in terms of competition: “tough competition has led to low standards” (2008, 486). I guess the idea is that neuroeconomists are competing with one another for fame and fortune, which forces them to rush to produce and publish results without taking the time and care required for rigorous elaboration and inference. Well, I think this might play a role, but I don’t think this is the whole explanation. Considerations of novelty and interdisciplinarity should also be in the story, as suggested earlier.

To have an even more complete story, I would like to add one more consideration. Neuroeconomics has certain qualities of “pop science” as it were. Pop science is that part of scientific work that addresses wider (academic or non-academic) audiences in contrast to expert audiences. One manifestation of this characteristic is that you can read reports of this work in newspapers and weekly magazines. That neuroscience has

characteristics of a pop science intended for lay audiences is understandable because people have a natural keen interest in themselves – in this case in the human brain. So neuroeconomics also has these qualities of a pop science, which means that it meets quite some extra-academic demand for insight, perhaps even for intellectual entertainment. Meeting this demand may create a pressure and opportunity for lowering the standards. Lower standards would not haunt the lay public as long as the image neuroeconomics as science is secured.

There is another possible external pressure that should not be ignored: the use of neuroeconomics for manipulation, such as in neuromarketing. There is a non-negligible risk of the results of neuroeconomics being used for social control that we might not like -- that would somehow violate what is valuable for us. So there are these external demands and there are the internal and intra-academic demands for scientific quality, and the two demands may be in conflict with one another. Neuroeconomics is shaped by these pressures and expectations that come from both outside and inside of disciplinary inquiry. The debate over the standards of quality has to be understood also in this context.

Let us finally briefly discuss some further questions of interdisciplinarity. Many advocates of neuroeconomics argue that economics and neuroscience can benefit from one another in both directions (e.g. Sanfey, Loewenstein, McClure, and Cohen 2008). This is a call for collaboration and mutual learning. Both disciplines are going to be the beneficiaries as soon as they start collaborating and learning from one another. What economics can offer for neurobiologists is a unified theoretical framework, lacking in neurobiology that is theoretically fragmented much as psychology is. So economics is believed to have this advantage that it can export to neurobiology. The justification of exportation here is epistemological and pragmatic. What neuroscience can offer is a multiple-systems approach, the idea being that there are multiple systems in the human brain that collaborate or compete with one another in shaping behaviour. The multiplicity of those systems inside the brain is of course absent in conventional economic theory. Economic theory has a different picture of the human mind, namely an image of one unitary all-purpose decision-making system resulting in rational choices. Many economists now suggest

that this picture be revised using the results of neuroscience. Here the justification is ontological, it is based on the idea that this is the way the human brain is.

The message from neuroscience is that human behaviour is a product of interaction of specialized subsystems in the brain. There is an automatic system (affective, heuristic-based System1) and there is the controlled system (deliberative, planful System2). The brain contains such different mechanisms that both collaborate and compete. The brain must coordinate the functioning of these systems with each other. This is where economics enters the picture as an exporter of theory – a theory of market coordination. So we get the idea of a tiny little market of neural activity in our brains, a market with the familiar capacity of coordination. The diverse and specialized neural systems are coordinated so as to result in coherent and purposeful behaviour. One of the exciting issues that this thought prompts is whether and what kind of ontological interpretation to apply when dealing with the metaphor of market coordination in the brain.

All of this has implications for boundaries of economics and the domain of rational choice. One question is, to what

kinds of behaviour does rational choice apply? One may have reasonably thought that deliberate rational choice must be fit for dealing with demanding and complex decision tasks with long-term tradeoffs; these are, after all, economically significant tasks. Based on experimental findings and models of brain structure, many now conclude that this is not the domain where rational choice is supported, since these decisions are largely governed by the affective mechanisms in the brain. “Ironically, rational choice models might therefore be most useful in thinking about the simplest kinds of decisions humans and other species make — involving perceptual tradeoffs, motor movements, foraging for food, and so forth—and prove least useful in thinking about abstract, complex, long-term tradeoffs which are the traditional province of economic theory.” (Camerer, Loewenstein, and Prelec 2005, 55) This would shrink the domain of rational choice to simple tasks of human decision-making (bad news to economics imperialism). A further step is to say its proper domain includes animal behaviour and neural activity in the brain (good news to economics imperialism).

Such drastic implications do not please everybody. Some economists try to devise arguments that would manage to

insulate conventional economics from the contaminating influence of contact with other disciplines. An extreme insulationist argument has been famously put forward by Gul and Pesendorfer (2008). They argue that neuroscience is in no way relevant to economics. Economists should simply ignore neuroscience, because the two disciplines address different levels of phenomena, those of choice and brain. Taking “revealed preference” or manifest choice to circumscribe the domain of economics, Gul and Pesendorfer set out to resist any attempt to use neurobiological data for casting doubt on the standard economic theory of rational choice: “Economic models make no predictions nor assumptions about body temperature, blood sugar levels, or other physiological data, and therefore, such data cannot refute economic models.” (2008, 19)

There are several exciting issues involved in this defensive argument, but let me just make one brief observation. To say that neurobiological data cannot refute economic models is probably correct. But it is not correct because there is no causal or constitutive connection whatsoever between the brain and choice behaviour. It is correct for the same reason for which no body of data is able to conclusively refute

economic models: models and theories are underdetermined by the data. This is of course a fact widely recognized in the philosophy of economics. The appropriate response to this is not to delimit the sources of evidence -- by excluding neurobiological data, for example -- but rather to broaden the range of such sources -- by including those data among many others. In case two or more economic models have the same or similar observational implications, it is advisable to consult whatever further evidence there may be available, then ascribe higher credibility to the neurobiologically sound ones - - those that are consistent with additional neurobiological data. This is what part of neuroeconomics does, and can be expected to do better in the future. It would be all too dogmatic -- an irrational choice -- to ignore its results as a matter of principle.

There are many more juicy issues prompted by the neuroeconomics initiative, and I expect them to proliferate in the near future. Regardless of how they will be resolved, the delightful fact that they are on the table and that they invite philosophical scrutiny is sufficient for satisfaction. But anything close to final judgement must wait.

BIBLIOGRAPHY

Camerer, Colin, George Loewenstein, and Drazen Prelec (2005) "Neuroeconomics: How Neuroscience Can Inform Economics", *Journal of Economic Literature*, XLIII, 9–64

Gul, Faruk and Pesendorfer, Wolfgang (2008) "The case for mindless economics", in *The Foundations of Positive and Normative Economics*, ed. by Andrew Caplin and Andrew Schotter. Oxford: Oxford University Press. pp. 3-39.

Mäki, Uskali (2009) "Economics imperialism: Concept and constraints", *Philosophy of the Social Sciences*

Rubinstein, Ariel (2008) "Comments on neuroeconomics", *Economics and Philosophy*, 24, 485-494.

Sanfey, Alan G., George Loewenstein, Samuel M. McClure, and Jonathan D. Cohen (2006) "Neuroeconomics: Cross-currents in research on decision-making", *Trends in Cognitive Sciences*, 10, 108-116.

ⁱ Más formalmente, si, para simplificar, nos limitamos a loterías de sólo dos resultados, pueden ser representadas de esta manera: $[x, p; y, (1-p)]$, en donde x e y representan los resultados o premios que pueden ser obtenidos si se las elige, y p y $(1 - p)$ las probabilidades asociadas a cada uno de ellos (en nuestro ejemplo, x podría ser la heladera y el microondas, y sería cero premio, y p tendría un valor de 0.2).

ⁱⁱ Véase Bernoulli, 1954.

ⁱⁱⁱ Frieman y Savage, 1952.

^{iv} Alchian (1953) ofrece una presentación sumamente accesible de este procedimiento. Véase también Friedman y Savage, 1948, p. 292.

^v Los fallidos más conocidos son las paradojas de Allais ("common" y "consequence effects") y Ellsberg (juicios probabilísticos inconsistentes), las "preference reversals" (Slovic y Lichtenstein), y los "framing effects" (Kahneman y Tversky).

^{vi} Machina (1982),

^{vii} Loomes y Sugden (1982).

^{viii} Las principales alternativas introducen transformaciones apropiadas en alguna de las variables clave de EUH, generando "fanning out", al operar sobre las utilidades (Machina, 1982) y transformando las probabilidades en "decision weights" (Kahneman y Tversky, 1979). Para una descripción de los cambios operados en EUH véase Schoemaker (1982) y Starmer (2000).

^{ix} Starmer (2000, pp. 360-363), agrupa a las desviaciones persistentes en tres ítems: a) violaciones de la monotonidad; b) event – splitting effects y c) violaciones de la transitividad.

^x "The basic principle of framing is the passive acceptance of the formulation given. Because of this passivity, people fail to construct a canonical representation for all extensionally equivalent descriptions of a state of affairs (...). they do not spontaneously transform the representations of puzzles or decisions problems. Obviously, no one is able to recognize '137 x 24' and '3288' as 'the same' number without going through some elaborate computations" (Kahneman (2003a, p. 1459).

^{xi} "Procedure invariance plays an essential role in measurement theories. For example, the ordering of objects with respect to mass can be established either by placing each object separately on a scale, or by placing both objects on the two sides of a pan balance; the two procedures yield the same ordering, within the limit of measurement errors. Analogously, the classical theory of preference assumes that each individual has a well defined preference order (or a utility function) that can be elicited either by offering a choice between options, or by observing their reservation price. Procedure invariance provides a test for the existence of a measurable attribute. It would have been difficult to attribute mass to objects if the orderings of these objects with respect to mass were dependent on the measuring device.

Similarly, it is difficult to defend the proposition that a person has a well-defined preference order (or equivalently a utility function) if different methods of elicitation give rise to different choices" (Tversky, 1999, p. 189; my italics).

^{xii} Friedman, 1955.

^{xiii} Para una defensa de la posibilidad de medir la utilidad en el sentido hedonista del término, véase Kahneman y Sugden, 2005.

^{xiv} Si en la consideración de un sujeto S, A es preferida a B, al elegir A el objetivo del sujeto es obtener del consumo ex – post de A la utilidad (experiencia hedonística) que esperaba ex – ante (o, al menos, mayor utilidad que la que obtendría de consumir B). El sujeto tiene una idea (una expectativa) de la utilidad que obtendrá de ese y otros bienes alternativos y es esta la que determina su elección. Esto significa que el sujeto prefiere A a B porque tiene una creencia (hipótesis) acerca del grado de satisfacción que le proporcionarán ambos bienes y puede compararlos a este respecto; y porque cree, además, que el consumo de A le brindará mayor satisfacción que el de B.

^{xv} Una consecuencia de lo dicho, que no desarrollaremos en este trabajo, es que el intervencionismo, al menos en su forma débil de "paternalismo", no sería fácilmente desechable en tales condiciones. El supuesto sustantivo de correspondencia es también importante, pues, porque se conecta con la defensa del no intervencionismo (laisse faire).

^{xvi} Al respecto véase Kahneman y Sugden (1985) y Kahneman y Thaler (1986).

^{xvii} La irracionalidad de estos intercambios no depende de que el sujeto pierda todo su dinero; basta con que emprenda una serie de transacciones que culmine con una disminución de su riqueza inicial.