

BREVE GENEALOGÍA DE LA MORAL POR ACUERDO¹

PEDRO FRANCÉS GÓMEZ
Universidad Complutense de Madrid

ABSTRACT

Neocontractarianism is probably the most conspicuous challenge to ethical utilitarianism. At the moment, Gauthier's *Morals by Agreement* can be considered the more advanced and developed version of moral contractarianism, as well as one of the most influential works about moral philosophy in recent years. Paradoxically, the threat it poses against utilitarianism takes its force from its deeply utilitarian roots. This essay tries to explain how the main contractarian concepts derived from quite simple ideas about rationality, prudence and morality, all of them having much to do with utilitarian conceptions. This is showed by tracing the main issues in Gauthier's major work back throughout his previous essays. As a result, the author maintains that a close link exists between utilitarian conceptions and the basis of contractarian thought, but precisely because of that, contractarian developments question the very utilitarian project as a whole. It is difficult to decide

¹ Este trabajo ha sido realizado en el marco de la investigación doctoral que desarrollo en el Departamento de Filosofía del Derecho, Moral y Política II (Ética y Sociología) de la Universidad Complutense de Madrid, bajo la dirección del Prof. Gilberto Gutiérrez López. Deseo hacer constar mi agradecimiento a las instituciones citadas, por su soporte económico y académico. Al Prof. Gutiérrez debo la orientación general de mi investigación, el apoyo constante, la idea de publicar estas páginas y diversos comentarios que han ayudado a mejorarlas; gracias por todo ello. También estoy agradecido por sus comentarios y ayuda al Prof. Carlos Thiebaut y a Julia García Maza, ambos del Instituto de Filosofía del C.S.I.C.

whether contractarianism has definitively overridden utilitarianism, but the suggestions offered by this genealogical study may shed some light on a pressing debate.

1. INTRODUCCIÓN

El neocontractualismo es en la actualidad el desafío más importante que tiene ante sí el utilitarismo. Tal vez sea un desafío definitivo; tanto más eficaz cuanto su fundamento teórico (así como su origen histórico y conceptual) es radicalmente utilitarista. Tal como se presenta en las conocidas y debatidas teorías de Rawls o Gauthier, el contractualismo se opone al utilitarismo sin negarlo; simplemente superándolo. En esencia, la tesis es la siguiente: Un grupo de agentes perfectamente racionales, en una situación ideal convenientemente caracterizada (una situación en la que se suprimiesen los factores *de hecho* que impiden a cada uno conocer con precisión y realizar aquellos actos encaminados a maximizar su utilidad) convendrían en someter sus acciones y sus instituciones a ciertos principios morales o de justicia, pues ése es el medio más adecuado para asegurar la promoción de sus intereses particulares, cualesquiera que éstos sean. Si esto es así —y el contractualista sostiene que, aceptadas ciertas ideas intuitivas simples sobre la racionalidad y el carácter de los agentes individuales, lo es— el argumento contractualista ofrece un criterio universal de validez de las normas morales, más abarcante y seguro que el criterio de la utilidad.

El contractualismo de Rawls presenta los principios de la justicia como alternativos a varios principios utilitaristas, y defiende que, dada su caracterización de la posición original y de las partes, los principios elegidos serían los de la justicia²; pero esa elección es realizada por individuos maximizadores, de los que no se niega que actúen correctamente al perseguir la máxima utilidad individual³. Se puede decir que el principio de utilidad está en la base del contractualismo rawlsiano, aunque no, desde luego, como principio moral. Sin embargo, la profunda relación entre utilidad y contrato aparece aún más clara en la obra de Gauthier, pues él,

² Cfr. John RAWLS, *Teoría de la justicia*, México, F.C.E., 1979, pp. 147 y ss.

³ Cfr. RAWLS, *op. cit.*, p. 41. Desde luego, aquí la maximización de utilidad no puede identificarse con una concepción «utilitarista» de los individuos en la posición originaria; pero sí creemos que está en relación con una idea utilitarista de la racionalidad individual.

desde una perspectiva liberal más radical, trata de fortalecer el nexo entre los intereses individuales concretos y la aceptación de principios morales; tal empresa le exige mayor profundidad en el análisis y descripción del carácter individual de las partes que contratan, así como la explicitación del tipo de racionalidad que poseen y de la historia o motivación personal que les conduciría a aceptar lo que él llama «restricciones morales racionales».

El resultado del análisis de Gauthier presenta, por un lado, individuos instrumentalmente racionales, gobernados por el principio de utilidad —utilitarismo que podemos calificar de «estándar», tomado de la economía—; y, por otro lado, una teoría ética que defiende que la moral no sólo es una necesidad racional sino que representa, además, un modo más elevado de racionalidad. Sin embargo, esta teoría ética apenas logra escapar a la crítica que la identifica con un artificio convencional construido por sujetos egoístas conscientes de que seguir fielmente ciertas reglas (morales) es el mejor medio para alcanzar sus fines a largo plazo. Según esta crítica, el contractualismo proporcionaría una justificación meramente instrumental para la moral. El contractualismo vendría a ser un utilitarismo de la regla extraordinariamente sofisticado⁴. No es así, desde luego, como los propios contractualistas ven sus teorías. Ellos consideran, por un lado, que la relación entre razón y moralidad es más profunda que la pura relación instrumental medios/fines (es decir, se niegan a aceptar que las reglas morales sean solamente el mejor medio posible para avanzar los intereses racionales a largo plazo de cada agente) y, por otro, que la dimensión moral puede ser legítimamente deducida de premisas no morales.

Sea de una forma o de otra; se trate de una versión sofisticada del utilitarismo de la regla o de un desafío definitivo a toda forma de utilitarismo, es del máximo interés obtener las claves teóricas de la forja de una idea más radical de las relaciones entre racionalidad y moralidad a partir de premisas utilitarias. Nosotros hemos creído que una breve historia o genealogía del pensamiento de Gauthier es un modo privilegiado de adentrarse en los vericuetos teórico-genéticos que permiten obtener esas claves. Una indagación sobre el origen de los conceptos que componen y caracterizan su teoría moral, justamente considerada como el

⁴ Tal vez más precisamente podría llamarse «utilitarismo de los principios», pues, según el contractualismo, son éstos, y no las reglas, los que han de someterse a la prueba de la utilidad, los que son elegidos según su utilidad.

más ambicioso intento de conciliar una razón puramente instrumental con una moral que responda al contenido de nuestras ideas previas e intuitivas sobre la misma⁵.

Esperamos que las páginas siguientes aclaren, al menos desde este punto de vista genético, pero también desde el argumental, la estructura elemental del contractualismo moral en su versión más depurada, y permitan una discusión más fructífera de sus premisas frente a las del utilitarismo. Se nos permitirá un método expositivo sencillo: comenzaremos con algún apunte biográfico, para ir después recorriendo los hitos más significativos en la empresa de Gauthier (la introducción de conceptos nuevos o los cambios en la orientación de su enfoque) en su orden cronológico. Así nos proponemos trazar el camino que, desde las primeras preocupaciones en los años sesenta, conduce a la teoría moral por acuerdo. La versión definitiva de la teoría, expuesta en *Morals by Agreement* iluminará nuestro trabajo, pero no será objeto de una atención sistemática.

2. LAS PREOCUPACIONES INICIALES Y LA INFLUENCIA DE HARE

David Gauthier inició los estudios de Filosofía en la Universidad de Toronto, su ciudad de nacimiento, y amplió su formación en Oxford, donde se doctoró en 1961. Su carrera académica se ha desarrollado en su mayor parte en el Departamento de Filosofía de la Universidad de Toronto, al que perteneció desde 1958 hasta 1980, año en que pasó al de la Universidad de Pittsburgh. El interés por los asuntos públicos (que le llevó en una ocasión a presentarse como candidato a diputado de la Cámara de los Comunes canadiense) y por la Filosofía Moral, estuvo presente desde el comienzo de su carrera: por un lado sus publicaciones no-académicas (columnas periodísticas, revistas de libros, etc.) trataron principalmente problemas relacionados con la política nacional y municipal y con los bienes públicos; por otro, su interés académico se centró en la filosofía moral de corte analítico (metaética, análisis del lenguaje moral, teorías de la argumentación moral) impulsada por Austin y Hare, de quienes fue discípulo en Oxford.

⁵ Nos referimos a la teoría moral contractual desarrollada principalmente en *La moral por acuerdo*, Barcelona, Gedisa, 1994.

David Gauthier comienza, por tanto, su investigación filosófica en un mundo anglosajón envuelto en las tradiciones analítica y utilitarista. Dentro de esa escuela, su influencia más notable provino, como él mismo reconoce, de Richard M. Hare, cuya obra *El lenguaje de la moral*⁶ lo estimuló a dedicarse a la investigación ética. Gauthier comprendió, según explica en el prefacio de su primera obra, *Practical Reasoning*, que la ética (por entonces condenada al emotivismo y al no-cognoscitivism, cuando no al irracionalismo) era un prometedor campo de estudio, dado que las características de la argumentación moral, y del fenómeno moral mismo, exigían que alguna forma de conexión entre moralidad y racionalidad fuese posible. De hecho, sus primeras obras se concentraron en el análisis de las relaciones entre racionalidad y moralidad, en un intento de conectar dos discursos que, desde la perspectiva analítica, parecían irremediabilmente alejados.

La necesidad de esta conexión había sido ya reconocida por Hare⁷ frente a las entonces influyentes tesis del existencialismo, el emotivismo y otras formas de no-cognoscitivism moral. Así, en la segunda página de *Freedom and Reason* podemos leer: «Contra esta convicción que todo adulto tiene de que es libre para formar sus propias opiniones sobre cuestiones morales, hemos de admitir otra característica de éstas que parece contradecirla. Se trata de que la respuesta a cuestiones morales es, o debería ser, una actividad racional». Y un poco más abajo añade: «La tarea de la filosofía moral, y la de este libro en particular, es buscar un camino para reconciliar estas posiciones aparentemente incompatibles, resolviendo así la antinomia entre libertad y razón». Hare creyó, en efecto, haber demostrado que los juicios morales, aunque reductibles a prescripciones, podían ser universalizables si se cumplían dos condiciones: primero, aceptar que tales juicios poseían cierto contenido descriptivo y, segundo, no considerar la universalidad como un principio moral sustantivo, sino como un principio lógico⁸. La universalizabilidad de los juicios morales garantizaba, según Hare, su racionalidad, pues los sometía a un principio lógico (de universalización) que él consideraba esencial-

⁶ México, Universidad Autónoma Nacional de México, 1975 (trad. de Genaro R. Garrido y Eduardo A. Rabosi). El original *The Language of Morals* se había publicado en 1952.

⁷ Véase especialmente su libro *Freedom and Reason* Oxford, Clarendon, 1963, pp. 2 y 3.

⁸ Véase Hare, R.M., *Freedom and Reason*, cap. 3, en especial el punto 2 (p. 31 y ss.).

mente idéntico al empleado en el lenguaje descriptivo, el cual era considerado paradigma de la racionalidad y no podía, por tanto, sino estar basado en principios racionales. Por otro lado, los juicios u opiniones morales, si bien autónomos, no son juicios caprichosos; han de estar basados en razones y ser coherentes con la regla de universalizabilidad. Así, sin negar la raíz autónoma del juicio moral, Hare creía contar con el criterio para apreciar su racionalidad.

3. LA PRIMERA OBRA: PRUDENCIA, MORALIDAD Y LENGUAJE

Sin embargo, el intento de reconciliación entre moral y razón llevado a cabo por Hare no fue considerado satisfactorio por Gauthier, quien dedicó su primer libro precisamente a una revisión de la relación entre el discurso prudencial (que toma como paradigma de la racionalidad) y el discurso moral. El largo título de aquella primera obra, *Practical Reasoning: The Structure and Foundations of Prudential and Moral Arguments and Their Exemplification in Discourse*⁹ (Oxford, Clarendon, 1963), anunciaba un análisis del razonamiento práctico en general (tanto prudencial como moral) que, sin duda por influencia de Kurt Baier y del propio Hare, tomaba como modelo el razonamiento prudencial, para defender después que la argumentación moral posee la misma estructura. Así, la relación que pretendía establecer Gauthier entre argumentación prudencial y argumentación moral era de identidad, aunque, evidentemente, ello no prejuzgara la relación entre la prudencia y la moralidad mismas. De todas formas, la conveniente identificación del discurso moral con el discurso prudencial permitía a Gauthier afirmar la racionalidad de la argumentación moral, que era lo que Hare había intentado sirviéndose de la regla de universalización.

La identificación de ambos discursos, prudencial y moral, necesaria para probar la racionalidad del último, se basa en la tesis de que ambos son formas del discurso práctico en general y en la formulación de una estructura simple para el argumento prudencial, en la que se determina qué elemento aporta «raciona-

⁹ «Razonamiento práctico: estructura y fundamentos de los argumentos prudencial y moral y su ejemplificación en el discurso».

lidad» al argumento. Después, el análisis del discurso moral revelará que posee la misma estructura ya definida para el prudencial, con lo que el elemento «racional» también está presente en él. Como hemos dicho, Gauthier defiende su tesis mediante el análisis de las argumentaciones prudenciales y morales. Los resultados de este análisis, que representan también lo esencial de su posición, merecen ser recordados, siquiera brevemente:

Los elementos principales de la estructura del argumento prudencial son: la condición del argumento, la conclusión y la regla analítica que liga a ambas. La condición del argumento es un deseo o necesidad (carencia) del agente. La conclusión será siempre una decisión. La deliberación que conduce desde el deseo o necesidad del agente hasta la decisión está guiada por una regla analítica de inferencia. El deseo o interés (primera premisa del argumento prudencial) puede muy bien ser irracional o, mejor, por depender exclusivamente de las preferencias subjetivas del agente, no es posible pronunciarse sobre su racionalidad «objetiva». Por otro lado, la decisión, en cuanto está ligada al deseo por una regla de inferencia invariable, posee su mismo carácter. Por lo tanto, es la regla misma de inferencia la que dota de racionalidad al argumento prudencial. Dicho de otro modo, es ella la que permitiría a cualquier otro ser racional anticipar la decisión de un determinado agente dados solamente sus deseos y las condiciones materiales disponibles. A continuación, analiza Gauthier el discurso moral y descubre que su diferencia con el prudencial no estriba en la regla de inferencia ni en la estructura del argumento (pues no es sino otra especie del discurso práctico en general). La única diferencia consiste en que la condición, en el caso del argumento que denominamos moral, es más amplia, de modo que abarca *todos* los deseos o necesidades *de todas las personas*. Pero si la peculiaridad del argumento moral reside sólo en el carácter de su condición y no en la regla de inferencia, no hay obstáculo en considerarlo tan racional como el argumento prudencial, con el que comparte el elemento lógico que lo dota de racionalidad.

La mayor originalidad del análisis de Gauthier consistió en negar que toda inferencia práctica hubiera de partir necesariamente de un principio general, pues define los deseos o necesidades del agente (sea un agente particular, en el caso de la prudencia o todos los agentes en el caso de la moral) como premisas particulares¹⁰. Tal concepción de la inferencia práctica le permitió

¹⁰ Cfr. *Practical Reasoning*, p. 169.

desarrollar la tesis que hemos resumido que, como puede apreciarse, supone una significativa reformulación de teoría de Hare sobre la relación entre lenguaje descriptivo y lenguaje prescriptivo.

Si creemos las palabras de Gauthier expresadas en aquel primer ensayo sobre una conexión entre prudencia y moralidad, éste obtuvo los siguientes resultados: criticó el principio de universalizabilidad como único criterio de racionalidad del juicio moral; aceptó y confirmó el carácter prescriptivo del lenguaje moral ordinario; y superó el «egoísmo ilustrado» de Baier al negar que el interés egoísta fuese la única premisa posible del argumento práctico¹¹. Sin embargo, no podemos olvidar que *Practical Reasoning* es una obra anclada en los métodos y en la orientación analíticos (como no podía ser de otra forma, al tratarse de un trabajo puramente académico¹², cuyo valor principal es servir como testimonio de que la inquietud de nuestro autor fue, desde sus inicios, la reconciliación de las formas divergentes de racionalidad que parecen presidir la argumentación prudencial y la argumentación moral; así como informarnos de que, también desde el comienzo, dio preeminencia al paradigma prudencial sobre el moral. Por lo demás, el análisis de Gauthier toma las inferencias prudenciales y morales como algo dado. La cuestión moral más radical, es decir, la cuestión de por qué tiene el agente prudente que razonar moralmente en ocasiones (que es la clave del problema de la relación entre racionalidad y moralidad si se concibe, como en el caso de Gauthier, que la argumentación prudencial es la argumentación práctica por antonomasia), no había sido tocada.

¹¹ De hecho, como hemos visto, el «interés de todas las personas» fue considerado también como una premisa de cierto tipo de argumento práctico, el argumento moral. Si entiendo bien la tesis de Gauthier frente a Baier, la diferencia sería la siguiente: el «interés de todos» es considerado por el segundo, como un punto de vista heurístico (el famoso «*moral point of view*») que el egoísta ilustrado adopta al tomar conciencia de que es lo que puede promover mejor «su verdadero interés». Por el contrario, Gauthier considera que para poder hablar siquiera de argumento moral en sentido propio, hay que *suponer* necesariamente que su condición es precisamente ésa: el interés o el deseo de todos. En caso contrario, no estamos ante un razonamiento moral. Para Gauthier, la premisa del interés de todos es definitoria del argumento moral; para Baier es parte del punto de vista al que conduce una teoría egoísta de la moral.

¹² Los diez primeros capítulos del libro (de un total de doce) son una versión de su tesis doctoral.

4. HOBBS Y LA TEORÍA DE JUEGOS: EL PUNTO DE PARTIDA

Tras la publicación de su primer libro, Gauthier se aparta un tanto del método analítico para iniciar un larga revisión de la historia del contractualismo, modelo teórico donde cree poder hallar el método adecuado para aportar su propia solución al problema de la conexión radical entre racionalidad prudencial y moralidad, mediante la construcción de una teoría contractual de la moral¹³. Paralelamente, publica varios artículos críticos sobre el utilitarismo de la regla y se interesa en un tema por entonces candente: el egoísmo como teoría ética. Fruto de estas investigaciones será la publicación, de modo sucesivo, de dos libros: el primero de ellos, de 1969, es *The Logic of Leviathan: The Moral and Political Theory of Thomas Hobbes*¹⁴, obra que pone de manifiesto la fascinación que el filósofo inglés ejerce sobre Gauthier (Hobbes será un tema recurrente en su bibliografía) y redescubre su filosofía política y moral, lo cual habría de tener repercusiones posteriores¹⁵. El segundo libro es una recopilación de artículos de varios autores, editado por Gauthier en 1970: *Morality and Rational Self-Interest* (Englewood Cliffs, Prentice Hall), donde se incluía el ensayo «Morality and Advantage» (del propio Gauthier)¹⁶. Ambos títulos son representativos de la nueva orientación de su pensamiento.

Merece especial atención el último ensayo mencionado: el artículo «La moral y la ventaja». Es un texto sencillo, cuyas tesis fundamentales habrían de ser reformuladas más tarde, pero posee el gran valor de ser el primer ensayo en que Gauthier emplea el Dilema del Prisionero¹⁷, así como rudimentos de Teoría de Juegos,

¹³ En el prefacio de *Morals by Agreement* (Oxford, Clarendon, 1986) Gauthier reconoce que la idea de una teoría contractual de la moral «capturó su imaginación» en el año 1966.

¹⁴ «La lógica de Leviathan: la teoría moral y política de Thomas Hobbes».

¹⁵ Nos referimos a la publicación, en 1986 de las dos interpretaciones contemporáneas de Hobbes más influyentes: las obras de Jean Hampton, *Hobbes and the Social Contract Tradition* (Cambridge, Cambridge U.P.) y Gregory Kavka, *Hobbesiana Moral and Political Theory* (Princeton, NJ, Princeton University Press). Ambas obras deben mucho a la rehabilitación de la filosofía moral hobbesiana emprendida por Gauthier.

¹⁶ Éste es uno de los pocos textos de Gauthier traducidos al castellano, se encuentra, bajo el título de «La moral y la ventaja» en J. Raz (ed.), *Razonamiento Práctico*, México, Fondo de Cultura Económica, 1986, pp. 341-363.

¹⁷ Para una introducción rápida, sencilla y sugerente del Dilema del Prisionero y su relevancia para la teoría de la racionalidad y la filosofía moral,

como herramienta analítica. En relación con esto, Gauthier declara, en el prefacio de *Morals by Agreement* (en lo sucesivo *MA*), que su investigación comenzó precisamente la tarde de noviembre de 1966 en que Howard Sobel le presentó el Dilema del Prisionero y algunas de las ideas básicas de la Teoría de Juegos¹⁸. Gauthier estaba ocupado, como ya hemos señalado, en encontrar un modo conveniente de formular la relación entre la moralidad y el interés. El lenguaje del contractualismo le parecía adecuado, pero ya había sido llevado a su máxima expresión por Hobbes; y la simple reinterpretación del *Leviatán* no resultaba satisfactoria. En ese punto, la Teoría de Juegos le ofreció no sólo un lenguaje especialmente idóneo para formular aquella relación, sino además un modo privilegiado de expresar el carácter problemático de la racionalidad prudencial y un medio técnicamente insuperable para desarrollar la teoría que tratara de solucionarlo.

En el artículo «La moral y la ventaja» distinguimos ya una formulación muy precisa de la conexión entre moral y beneficio individual que Gauthier está buscando. Así, reconoce que el papel de la moral consiste en limitar la búsqueda individual del beneficio cuando ese límite sirve a un interés mayor de todos. Su visión de la moral es deudora aún de la de Kurt Baier, pero muestra ya incomodidad respecto a la misma, prefigurando en ciertos pasajes su propia concepción. Por otro lado, el análisis de la deliberación prudencial se basa en la Teoría de Juegos, y la pregunta que quedaba obviada en *Practical Reasoning* aparece explícita: Gauthier parte ya, en su análisis, del individuo prudente que se plantea directamente esta pregunta «¿Por qué debo ser moral?». Ciertamente, el artículo revela que la relación entre prudencia y mora-

puede verse el breve artículo de Derek Parfit "Prudencia, moralidad y el dilema del prisionero", en *Diálogo filosófico*, 13 (1989), pp. 4-30.

¹⁸ La Teoría de Juegos fue formulada en 1944 por J. von Neuman y O. Morgenstern (*Theory of Games and Economic Behavior* Princeton, Princeton U. P., 1944) con el fin de proporcionar un modelo explicativo de ciertos comportamientos económicos caracterizados por la presencia de dos o más agentes cuyas respectivas decisiones pueden dar lugar a distintos resultados según las decisiones de los demás. La teoría de juegos ofrece un medio para analizar cualquier contexto de interacción estratégica, es decir, contextos en los que las decisiones que uno toma han de responder a las decisiones que espera que los demás hayan tomado, y a la inversa, las decisiones de cada agente influyen en las decisiones de los demás. La teoría fue desarrollada, entre otros, por J.F. Nash, R.D. Luce y H. Raiffa. En concreto, el libro de Luce y Raiffa, *Games and Decisions* (Nueva York, John Wiley & Sons, 1957) supuso la generalización de la teoría y el inicio de su uso, como parte de la teoría de la Decisión Racional, fuera del estricto ámbito de la economía.

lidad sólo era entonces, para Gauthier, un problema: la solución teórica ni siquiera era aún vislumbrada, ya que aquel pequeño ensayo concluía precisamente desconfiando de que un agente prudente (y sólo prudente) pudiera llegar a abrazar sinceramente el sistema de restricciones morales, con el coste neto que éste puede llegar a suponer desde la perspectiva del interés individual.

«La moral y la ventaja» evidencia que, aunque Gauthier se había desprendido por completo de sus orígenes analíticos para profundizar en el estudio de la racionalidad como auto-interés y su relación con la moralidad, todavía no había dado forma definitiva a su concepción de la racionalidad económica ni había transformado su concepción de la moralidad en la línea «rawlsiana» en que se movería después. Su proyecto podría definirse, en aquel momento, como el intento de conciliar el egoísmo racional con la moralidad. Y era un proyecto que le conducía más bien a tener que reconocer, a su pesar, que o bien no había lugar para la moral entre egoístas racionales, o bien la moral que puede existir entre ellos no se ajusta a lo que intuitivamente definimos como tal.

5. EL IMPACTO DE LA *TEORÍA DE LA JUSTICIA* DE RAWLS

Sin embargo, tal proyecto habría de ser decisivamente reformulado e impulsado tras, primero, la profundización en la Teoría de la Decisión Racional y la Teoría de Juegos, que sólo aparecen de forma rudimentaria en «La moral y la ventaja» y, segundo, la aparición de *La teoría de la justicia* de John Rawls. El empleo de la Teoría de la Decisión Racional le permitió, por ejemplo, la potente formulación de la tesis de la imposibilidad del egoísmo racional¹⁹, con la que se deshacía de uno de los fantasmas que le había perseguido en los años anteriores. Por su lado, la aparición de la obra de Rawls, sirvió como punto de referencia polémico para perfilar los contornos de la teoría neocontractualista de Gauthier.

De hecho, la mayoría de edad de su proyecto puede datarse en 1974, cuando presenta su «Justice and Natural Endowment: Toward a Critique of Rawls's Ideological Framework»²⁰ («Justicia

¹⁹ Esta tesis está recogida en su antológico artículo «The Impossibility of Rational Egoism», *Journal of Philosophy*, 71 (1974) 439-456.

²⁰ En *Social Theory and Practice*, 3, 1975, 3-26. Recogido en Gauthier, *Moral Dealing*, Ithaca, Cornell U.P., 1990, pp. 150-170. Para las referencias empleamos esta segunda edición.

y dotación natural: hacia una crítica del marco ideológico de Rawls»). Esta contribución es bien conocida entre los comentaristas de Rawls por su sólida crítica de la deducción rawlsiana de los principios de la justicia a partir de la posición original, así como por la construcción de una versión alternativa de la misma, que desde entonces ha sido identificada como *la alternativa «liberal»* del contractualismo moral. No tematizaremos aquí las profundas diferencias entre el contractualismo de Rawls y el de Gauthier, que son precisamente uno de los acicates de la obra del segundo. Nos limitaremos a decir que Gauthier ofreció en 1974 el esquema de una teoría contractualista de la justicia que, partiendo de los presupuestos aceptados por Rawls²¹, conducía, no a los conocidos principios defendidos en la *Teoría de la Justicia*, sino a los principios retributivos que rigen el intercambio de mercado.

El razonamiento de Gauthier niega que, desde la ideología liberal que Rawls asume, y mediante una teoría contractualista de la sociedad, se deduzca el segundo principio de la justicia (el principio lexicográfico de la diferencia) tal como Rawls lo concibe. Si tal principio hubiera de regir la distribución de bienes sociales primarios, pronto aparecería con claridad que, en esa distribución, las capacidades y talentos naturales de los miembros de la sociedad se toman como si fuesen propiedad común, pues no juegan ningún papel en la distribución determinada por los principios de la justicia. Pero ello contradice el individualismo y la racionalidad maximizadora que Rawls sitúa en la base de la deducción de los principios. Frente a esa contradicción, Gauthier expone lo que según él es la verdadera teoría contractualista que surge de los presupuestos liberales de Rawls. Esta teoría asumiría con seriedad el carácter maximizador e individualista de la racionalidad, y replantearía, desde tal asunción, el acuerdo que es racional adoptar en la posición original. El resultado es que, aun concediendo que el acuerdo sobre los principios de la justicia debe tomarse «tras el velo de ignorancia», los agentes en la posición original, conscientes de que entre ellos existen diferencias naturales (aunque no sepan si ellos particularmente serán o no favorecidos), no acordarán principios que ignoren estas diferencias, sino, al contrario, principios que, aun garantizando el beneficio mutuo que es la base de la cooperación, aseguren también la justa retri-

²¹ Los presupuestos de Rawls, tal como son identificados por Gauthier en su ensayo, corresponden a los presupuestos de una sociedad liberal: el individualismo y la concepción maximizadora de la racionalidad.

bución a cada uno según su aportación a la empresa común que es la sociedad.

La defensa del contractualismo liberal frente al «equivocado» contractualismo igualitarista de Rawls incluye críticas y reducciones al absurdo de algunos de los pasos de la deducción rawlsiana de los principios, pero también exige que Gauthier ofrezca una explicación positiva de lo que, según él, es la deducción correcta. Y tal aclaración exige, a su vez, la definición de conceptos cuya importancia reside en que prefiguran algunos de los conceptos centrales de la teoría moral por acuerdo que habría de aparecer doce años después. Quizá sea pertinente poner sobre el tapete estos conceptos, aunque debamos, para ello, profundizar algo más en el argumento que Gauthier emplea para criticar la teoría de Rawls.

Gauthier no acepta el principio de la diferencia en su versión rawlsiana porque cree que los individuos que lo acordaron en la posición original no seguirían viéndolo como un principio de justicia racional, una vez levantado el velo de ignorancia. Gauthier piensa (y en este punto coincide con Rawls) que la posición original ha de ser una perspectiva que cualquier miembro de la sociedad debe poder recuperar en cualquier momento, para confirmar la racionalidad de los principios de la justicia. Pues bien, el principio de la diferencia no resiste tal confirmación retrospectiva. Una vez levantado el velo, sabedores de sus talentos naturales, cada agente auto-interesado puede preguntarse si no habría sido más racional (en el sentido de maximizador para las expectativas de cada miembro de la sociedad) reflejar esa diferencia natural en la estructura social. Gauthier sostiene que el único modo de evitar esta sospecha retrospectiva sería incluir en la posición original la conciencia de que las diferencias existen, y de que, en ausencia de cooperación, cada agente podría esperar cierto beneficio, derivado solamente del comportamiento egoísta de todos y de su dotación natural. Ningún agente racional renunciará a ese beneficio no-cooperativo; ni admitirá una estructura social en la que se vea privado de parte de él. Por tanto, el principio de la diferencia, que determina la distribución de bienes primarios, regirá, en efecto, la distribución de los bienes sociales, pero sólo de aquellos que sean productos específicos de la cooperación social, no de aquellos otros que ya cada individuo habría adquirido en ausencia de la cooperación social. Así, lo que está en juego es lo que Gauthier denomina «el *excedente social*», es decir, la parte de bienes primarios que se deben al hecho de la cooperación, excluyen-

do la parte de bienes que de todas formas se habrían producido en ausencia de cooperación. Estos últimos ya estarían «naturalmente» distribuidos según los talentos y capacidades de cada agente, de modo que un principio justo de cooperación (cuya misión es asegurar la cooperación mutuamente ventajosa mediante la distribución «imparcial» de los beneficios de la misma) tiene, primero, que reproducir la que habría sido una distribución natural, y sólo después intentar la imparcialidad en la distribución del excedente cooperativo; de otro modo, sería visto como injusto por quienes más contribuyen y más se esfuerzan en beneficio de todos.

Ahora bien, una vez establecido que el principio de la diferencia debe aplicarse al excedente social y no al conjunto total de los bienes básicos (con lo que se ha introducido una severa modificación «liberal» en la teoría de Rawls), cabe preguntarse si la regla de distribución que incorpora ese principio es adecuada o no. En principio la regla no es inadecuada, en el sentido de que ordena distribuir los bienes sociales primarios de modo que cualquier desigualdad suponga mayor beneficio de los menos aventajados. Tal regla parece aceptable por todos en la posición original, desde la conocida perspectiva de la maximización del beneficio que cada uno espera obtener de la cooperación, unida a la ignorancia sobre el lugar que cada uno ocupará en la sociedad. Sin embargo, es una regla ambigua, porque existen al menos dos posibilidades para medir el beneficio que cada uno espera obtener: cabe considerar que el beneficio es paralelo a la *cantidad* de bienes primarios, pero también cabe medirlo según cierta *proporción*: según la medida en que los bienes contribuyen al bienestar particular de cada uno. Una vez más, la opción de Rawls es la primera, mientras que la opción más adecuada conforme a la racionalidad maximizadora que Rawls escogió como punto de partida, es la segunda. En efecto, el beneficio que proporciona un bien primario no es directamente función de la cantidad de bienes, sino de la *utilidad* de esos bienes, y la utilidad dependerá de las expectativas que cada individuo tenga. Cada individuo, dada su dotación natural, tiene una *expectativa social máxima*²², cuya satisfacción completa representaría el mayor beneficio que podría obtener de la cooperación. Cuanto menor sea la proporción en que esa expectativa queda satisfecha, menor beneficio obtiene el agente. Por tanto, el beneficio no depende de la canti-

²² Gauthier denomina a esta expectativa el «potencial social», al referirse al beneficio potencial que el agente puede obtener de la sociedad; cfr. «Justice and Natural Endowment...», p. 162.

dad de bienes primarios, sino de la proporción en que cada individuo ve cumplida su expectativa máxima. Así pues, el “principio proporcional de la diferencia” (como lo denomina Gauthier) debe leerse así: la distribución del excedente social debe ser tal que todos vean su expectativa máxima cumplida en la misma proporción, salvo que cierta desigualdad sirva para que el menos favorecido vea aumentada la proporción en que se satisface su expectativa.

Según Gauthier, el principio proporcional de la diferencia sería el realmente acordado por sujetos maximizadores en una posición original como la descrita en *La Teoría de la Justicia*. Y tal principio no da lugar a la unión social cuasi-igualitaria que Rawls defiende, sino a una sociedad semejante a la sociedad capitalista de mercado, «porque el mercado competitivo es el mecanismo por medio del cual se produce y distribuye el excedente social óptimo de acuerdo a la contribución de cada persona»²³.

Gauthier concluye su crítica a Rawls con pesimismo: el desarrollo de una teoría contractualista basada en nuestra intuición sobre lo que es la racionalidad produce como resultado un principio de justicia que choca con nuestra idea intuitiva de la moral. Así pues, o bien la concepción de racionalidad más extendida entre los teóricos sociales es errónea, o nuestra idea ordinaria de la moral está equivocada. Corregir esa idea ordinaria choca tan frontalmente con la tradición filosófico-moral que ningún filósofo moral estaría dispuesto a hacerlo —dice Gauthier. Pero mantener nuestra visión ordinaria de la moral supone aceptar que es irracional, al menos desde el punto de vista adoptado. A menos que una nueva formulación de la racionalidad, capaz de superar los estrechos límites de la concepción maximizadora, entre en escena y resulte tan plausible como aquélla. Gauthier intentará, precisamente, aportar esa nueva formulación en sus siguientes ensayos, pero antes de exponer ese intento, conviene recapitular el significado global de la crítica de Gauthier a Rawls.

Como hemos visto, Gauthier se enfrenta al contractualismo de Rawls sin renunciar al paradigma contractualista, pero afirmando de un modo más radical el origen liberal del mismo. Incluso se percibe cierto legado lockeano²⁴ en conceptos tales como dotación

²³ «Justice and Natural Endowment...», p. 169.

²⁴ Aunque en modo alguno debe confundirse la teoría definitiva de Gauthier con una teoría retributiva de la justicia, del tipo de la de Nozick. A continuación tematizaremos algunos de los elementos que evidencian la distancia

natural, beneficio esperado en ausencia de cooperación, etc. Con ello, Gauthier se sitúa desde entonces como defensor de un modelo liberal de contractualismo que incluye entre sus conceptos básicos los de «excedente social» y «expectativa máxima», que aparecen ya en el artículo que comentamos²⁵. Lo más interesante es que en este ensayo Gauthier da forma a una versión del contrato muy diferente de la de Rawls, bien fundada y argumentalmente impecable, pero sólo para mostrar la imposibilidad de llegar a una conclusión moral convincente a partir de premisas estrictamente liberales. Tal vez por eso, Gauthier necesitaba revisar la principal de ellas: el estrecho concepto de racionalidad como maximización.

6. EL CONCEPTO DE «RACIONALIDAD RESTRINGIDA»

Gauthier dedicó grandes esfuerzos a desarrollar su hoy bien conocida, aunque polémica, teoría de la racionalidad como maximización *restringida*, que fue su respuesta al dilema con el que concluye «Natural Endowment...». Esta teoría apareció por primera vez en su artículo «Reason and Maximization» («Razón y maximización»)²⁶, y estaba destinada a jugar un papel esencial en su teoría moral posterior. De hecho, es probablemente la concepción clave —y una de las mayores originalidades— de su obra. La versión definitiva del concepto de racionalidad restringida se encuentra en el capítulo VI de *MA*²⁷ y en algunos textos posteriores²⁸; sin embargo, nos centraremos ahora en algunos rasgos de su primera formulación, tal como corresponde al objetivo «genealógico» de este trabajo.

entre Gauthier y Nozick; sin embargo, la explicitación de las diferencias entre ambos modelos contractualistas escapa a la dimensión de este artículo.

²⁵ Aunque, como me ha sugerido el prof. Thiebaut, no es legítimo desposeer a Rawls y a otros teóricos del «título» de liberal (lo que nos llevaría a tener que precisar mucho en qué sentido lo es Gauthier), queremos resaltar el carácter diferencial de Gauthier, especialmente por los resultados finales de su teoría (más liberales en la práctica), para reflejar su propia opinión sobre la visión rawlsiana.

²⁶ *Canadian Journal of Philosophy*, 4, 1975, pp. 411-433. Reimpreso en *Moral Dealing*, pp. 209-233. Emplearemos esta última edición en las referencias.

²⁷ *Morals by Agreement*, Oxford, Oxford University Press, 1986, pp. 157-190. En la traducción española: *La moral por acuerdo*, Barcelona, Gedisa, 1994, pp. 213-255.

²⁸ Se trata de reflexiones posteriores a *MA*, como las contenidas en «Rational Constraint: Some last Words», en Vallentyne, P. (ed.), *Contractarianism and Rational Choice*, Nueva York, Cambridge U.P., 1991, pp. 323-330; «Moral Arti-

La maximización restringida (Constrained Maximization) es concebida, en el artículo de 1975 «Reason and Maximization», como un principio general de la racionalidad, alternativo al principio (o condición) de la racionalidad *directamente maximizadora*. Por racionalidad directamente maximizadora entenderemos, desde ahora, el modelo de racionalidad individual ampliamente aceptado en las ciencias sociales (Economía, Sociología, Psicología, Teoría de la Decisión, Teoría Política, etc.), basado, por un lado, en la Teoría Bayesiana de la Decisión y la Teoría de Juegos y, por otro, en la definición matemática de utilidad como medida o expresión de las preferencias individuales. Este modelo postula como condición de la racionalidad el siguiente principio: «Una persona actúa racionalmente sólo si el resultado esperado de su acción le proporciona una utilidad al menos tan grande como el resultado esperado de cualquier otra acción posible para él en la situación dada»²⁹. Este principio exige que un agente racional trate de «maximizar» su utilidad, es decir, elija o decida siempre de tal modo que pueda esperarse de su acción, —según la información disponible, las condiciones de hecho y/o las acciones esperadas de los otros— que satisfaga sus preferencias (expresadas en su función de utilidad) en la *mayor medida* posible. Un breve ejemplo³⁰ puede aclarar el funcionamiento normativo del principio: Imaginemos que un agente se topa con dos billetes de banco, uno de mil y otro de dos mil pesetas, y alguien le dice que le está permitido apropiarse de uno de ellos, pero sólo de uno. En esta situación, y suponiendo primero, que lo único que está en juego es el valor de cambio del dinero que representan los billetes (es decir, eliminando posibles preferencias estéticas o de otra índole); segundo, que el valor del dinero representa utilidad en general para cualquier agente (más dinero implica más cantidad de aquello que el agente prefiera, o la misma cantidad durante más tiem-

», *Canadian Journal of Philosophy*, 18 (1988), pp. 385-418, y especialmente en «Assure and Threaten», *Ethics*, 104 (julio 1994), pp. 690-721.

²⁹ Así aparece enunciado en Gauthier, «Reason and Maximization», p. 216.

³⁰ Para ejemplificar o mostrar exactamente la racionalidad de una elección, es necesario estar seguros de que nos hallamos ante una decisión en estado puro. En las decisiones de la vida cotidiana, es frecuente que las utilidades de las personas se influyan mutuamente (existe la simpatía y el altruismo), también ocurre que las escalas de preferencias de un mismo individuo se entrecruzan (se prefiere más dinero, pero porque «representa» más prestigio, o más poder). Estas «impurezas» de las decisiones reales impedirían el análisis que la Teoría de la Decisión exige; por eso el ejemplo, aún tomando pie en la realidad e intuitivamente ilustrativo, es una situación ficticia que ha de cumplir perfectamente ciertas condiciones hipotéticas.

po, etc.); y, tercero, que la elección del agente no tiene ninguna influencia sobre las utilidades de agentes no implicados, ni sobre utilidades para él mismo fuera de la mera utilidad proporcionada por el valor monetario (es decir, el dinero que no recoja no podrá ser encontrado por nadie más, ni el hecho de tomar uno u otro billete le reportará mejor o peor fama: no hay valoraciones morales asociadas a la elección), por lo que se trata de una decisión puramente prudente. En esta situación, decíamos, es racional la conducta de quien tome el billete de más valor. Tomar el billete de mil pesetas habría contradicho el principio de la racionalidad maximizadora, pues el resultado de esa acción (tomar el billete de mil pesetas) le proporciona menos utilidad que la acción alternativa disponible en ese momento. Mientras, esta segunda acción se adecúa al principio: no hay ninguna otra, en las circunstancias del caso, que le proporcione mayor utilidad.

Este modelo de racionalidad es asumido tácitamente por la mayoría de científicos sociales (especialmente por economistas y teóricos de la política y la elección pública) cuando realizan predicciones sobre cómo actuará un individuo o un grupo; y es explícitamente defendido por la tradición filosófica empirista y utilitarista.

Sobre la racionalidad como maximización deben hacerse dos advertencias inmediatamente, aunque tendremos ocasión de profundizar en ellas más adelante. La primera es que el concepto mismo de «maximización de la utilidad» se basa en la posibilidad de definir una función matemática de utilidad a partir de unas preferencias individuales dadas. Ello exige que las preferencias individuales cumplan ciertas condiciones elementales de coherencia, en las que no nos detendremos ahora. Quiere esto decir que la propia definición de racionalidad como maximización de la utilidad presupone que las preferencias del agente poseen cierta estructura y que el propio agente es capaz de ordenarlas de un modo coherente. Estas características de las preferencias y del agente forman también parte de la racionalidad como maximización, aunque no aparezcan explícitamente en su principio, pues están supuestas por la idea de utilidad. Una segunda advertencia se refiere al *contenido* de las preferencias. Éstas son, en efecto, preferencias *del* agente, pero ello no implica necesariamente que sean preferencias *referidas al* agente. La racionalidad como maximización no se preocupa por el contenido de las preferencias. Un acto se califica como racional si maximiza la utilidad de cierto agente, pero las preferencias de ese agente se toman como dadas.

Y pueden muy bien ser preferencias del todo altruistas. Es decir, no hay que confundir una concepción individualista de la racionalidad, con una concepción egoísta.

Esta breve caracterización de la racionalidad maximizadora nos permite ya acercarnos a la crítica que Gauthier formula contra ella.

La racionalidad como maximización parece funcionar perfectamente cuando los agentes actúan independientemente, es decir, cuando en su decisión sólo han de tener en cuenta los posibles resultados de *sus* acciones en relación con circunstancias fijas (sean éstas circunstancias conocidas con certeza, o sólo con alguna probabilidad, o incluso totalmente desconocidas). En estos casos, el agente que se conduzca según el principio de la racionalidad maximizadora podrá tener la seguridad de que no hay otra vía de acción más prudente que la elegida por él: obtendrá tanto beneficio como sea posible obtener en las circunstancias en que se encuentre. Sin embargo, en ciertas situaciones de interacción con otros agentes igualmente racionales, el resultado de las acciones maximizadoras de cada uno de ellos, puede proporcionar a cada uno una utilidad *menor* que la que podrían haber obtenido ambos siguiendo un principio diferente. En estos casos, ejemplificados por el Dilema del Prisionero (al que aludíamos arriba), la racionalidad maximizadora no logra su objetivo de guiar la acción hacia la obtención de la mayor utilidad posible. O, por decirlo con más precisión, lo logra pírricamente, a costa de prohibir a los individuos emplear estrategias aún más beneficiosas para ellos que la propuesta por el principio de la racionalidad maximizadora individual. Explicaremos esto con más detalle: en las situaciones del tipo Dilema del Prisionero, cada individuo actúa racionalmente desde su punto de vista individual, pero un observador imparcial comprueba que *ambos* jugadores podrían obtener más utilidad si en vez de seguir una estrategia maximizadora optasen (ambos) por una estrategia diferente. Además, cada jugador *es consciente* de ello pues, como agente racional, ha de poder figurarse la situación completa. De manera que en estas situaciones se da la bien conocida paradoja de que la racionalidad maximizadora resulta impotente para realizar justamente lo que ella recomienda, esto es, alcanzar la mayor utilidad posible en cada situación dada.

El problema aquí planteado puede generalizarse introduciendo los conceptos de equilibrio y optimalidad. Tratar de explicar aquí completamente estos conceptos nos llevaría demasiado lejos; simplemente diremos que en los casos representados por el Dile-

ma del Prisionero, se produce un resultado en equilibrio (ninguno de los jugadores tiene razones para elegir una acción diferente), pero no óptimo (hay resultados alternativos y factibles que proporcionarían más utilidad a alguna de las partes sin disminuir la utilidad de nadie; en este caso, el incremento de utilidad sería para todos). Cualquier individuo racional está interesado en que el resultado de su interacción sea siempre óptimo. La optimalidad es lo mínimo a lo que un agente maximizador puede aspirar. Después, dado un resultado óptimo, cada agente pretenderá que su «pago» (la utilidad que el juego le asigna como resultado) sea el mayor posible; pero, por lo menos, preferirá que no se «desperdicie» utilidad obteniendo un resultado sub-óptimo, cuando el óptimo es factible. Alcanzar la optimalidad parece ser una exigencia de la racionalidad (aun entendida como simple maximización).

Sin embargo, como veíamos, el lenguaje del principio de la racionalidad no incluye referencia alguna a la optimalidad. La optimalidad representa un punto de vista global que, a la larga, proporciona mayor utilidad a cada agente, pero que cae fuera del punto de vista de la maximización directa de utilidad individual. Ello es así porque el alcanzar un resultado óptimo exige que todos los agentes actúen conforme a unas estrategias prefijadas, que pueden contradecir a la estrategia maximizadora. De este modo, la persecución de la optimización exige un principio de actuación racional diferente. Este principio es formulado por Gauthier en la p. 227 de «Reason and Maximization», y dice así:

«Una persona que actúa interdependientemente actúa racionalmente sólo si el resultado esperado de su acción le proporciona a cada una de las personas que interactúan una utilidad tal que no haya ninguna combinación de acciones posibles (una para cada persona de las que interactúan, incluyendo al agente) con un resultado esperado que proporcione a las demás personas una utilidad al menos igual, y a ella misma una utilidad mayor».

Se trata de un principio optimizador porque, en caso de interacción, exige al individuo racional que no tenga en cuenta las acciones de los otros, sino sus utilidades, y actúe conforme a la combinación de acciones que produzca un resultado óptimo. El principio no supone la maximización como se entiende normalmente en la Teoría de la Decisión. Exige un comportamiento maximizador en caso de acción independiente, y un comportamiento optimizador en caso de acción interdependiente. Con esta política, el agente no pierde utilidad en las primeras situaciones (respecto al maximizador directo, seguidor del principio de la

racionalidad «tradicional») mientras gana mucho en las del segundo tipo (pues logra beneficios vedados al maximizador directo). Así, en conjunto, obtiene una utilidad mayor.

Con esto, Gauthier ha presentado un principio alternativo de la racionalidad. Lo que hace a continuación es bautizarlo y, posteriormente, defenderlo frente al principio de la racionalidad simplemente maximizadora. En cuanto al bautizo del principio, dice:

«llamaré a ésta la condición de la *maximización restringida* (*constrained maximization*). Y por maximización restringida entiendo esa política, o cualquier política, que requiere la maximización de utilidad individual en el estado de naturaleza y una optimización acordada en sociedad»³¹.

En cuanto a su defensa, exige la introducción de algunas hipótesis y el desarrollo de un argumento complejo que resumiremos diciendo que, si ambos principios —el de la maximización directa y el de la maximización restringida— se someten a la prueba de ser ofrecidos a un hombre económico como posibles modelos de racionalidad a adoptar, éste elegirá el segundo de ellos. Es decir, que un maximizador directo, con criterios simplemente maximizadores, *preferiría*, si pudiera elegir, convertirse en maximizador restringido. Un maximizador restringido decidiría, por cierto, lo mismo. En definitiva, Gauthier concluye que su modelo de racionalidad restringida es autofundante, mientras la racionalidad simplemente maximizadora no lo es³².

Esta conclusión abre nuevas expectativas para el hombre económico quien, provisto con esta nueva interpretación de la racionalidad, es capaz de superponer a la mera maximización directa, una revisión de su estrategia que tiene mucho parecido con un punto de vista moral. En efecto, aunque este aspecto es muy discutible, Gauthier considera que la racionalidad restringida equivale a cierta interpretación de la moralidad. Con ello, la moralidad (al menos así entendida) entraría a formar parte de las

³¹ «Reason and Maximization», Cit., pp. 277-278.

³² Pues un maximizador directo elegiría, guiado por su criterio de racionalidad, un tipo de racionalidad diferente; o, desde otro punto de vista, su modo de actuación teóricamente racional resulta ser irracional a la luz de su propio principio. Estas consideraciones sobre la racionalidad como maximización llevaron a Derek Parfit (*Reasons and Persons*, Oxford, Oxford U.P., 1986, cap. 1) a calificarla como «teoría auto-frustrante» (*self-defeating*). La racionalidad restringida resulta racional según sus principios, por lo que escapa a la crítica de Parfit y es considerada «autofundante» por Gauthier.

características del *homo oeconomicus* o simplemente prudente, aunque, como resalta Gauthier, de un modo paradójico:

«Podríamos expresar la relación entre prudencia y moralidad precisa, aunque paradójicamente, diciendo que el hombre prudente considera racional *hacerse* moral, pero no *serlo*. El hombre prudente puede justificar sobre bases prudenciales la adopción de un fundamento moral (en vez de prudencial) para la acción; pero sólo una vez adoptada esa base moral, sólo una vez convertido en un ser moral, puede justificar un plan de acción moral, en vez de prudencial»³³.

7. RACIONALIDAD ECONÓMICA Y MORALIDAD

Se puede decir que la visión de la racionalidad y su relación con la moralidad quedaba fijada en el artículo «Reason and Maximization». Ahora bien, lo que nos ofrece «Reason and Maximization» es un boceto que, por un lado, habría de ser concluido más adelante y, por otro, representa sólo un fragmento de un plan mayor, que se completa con otros dos artículos relevantes sobre el concepto de racionalidad, también de los años setenta. El primero de ellos es «Rational Cooperation» («Cooperación racional») ³⁴, en el que se había puesto de manifiesto la necesidad de un procedimiento racional de negociación para alcanzar acuerdos sobre los conjuntos de estrategias que podrían conducir a resultados sociales óptimos. «Rational Cooperation» complementa el punto de «Reason and Maximization» en que, reconociendo que es más beneficioso en términos maximizadores adoptar una estrategia optimizadora, hay que elegir conjuntamente cuál de las posibles estrategias optimizadoras se pone de hecho en práctica. Tal elección demanda una solución racional, pero Gauthier, en este momento, sólo puede decir que será un arduo problema de negociación entre las partes.

El segundo de los artículos que completa la visión de la racionalidad bosquejada en «Reason and Maximization» es «Economic Rationality and Moral Constraints» ³⁵ («Racionalidad económica y restricciones morales»). Se trata de un texto que ya ha profundi-

³³ *Ibid.*, pp. 232-233.

³⁴ En *Noûs*, 8 (1974), pp. 53-65.

³⁵ En *Midwest Studies in Philosophy*, 3 (1978), pp. 75-96.

zando tanto en la caracterización matemática de la racionalidad económica como en la concepción sobre la relación entre racionalidad y moralidad. En él se contiene, además, una defensa (la primera, y una de las más decididas) de su proyecto contractualista moral. La defensa consiste en contrastar las teorías éticas más importantes con la racionalidad económica; es decir, poner a prueba la «racionalidad económica» de las éticas más difundidas (utilitarismo, ética de los derechos naturales y contractualismo). La racionalidad económica se ha asociado siempre al utilitarismo ético, por lo que el resultado esperado del test heurístico que Gauthier realiza sería precisamente la confirmación de que la racionalidad económica apoya una ética utilitarista (como defiende Harsanyi, por ejemplo, frente a Rawls). Sin embargo, Gauthier muestra cómo la racionalidad económica se compadece mejor con una moral contractual. Ésta recoge el espíritu del hombre económico, a la vez que amplía su horizonte de modo inesperado.

Con el artículo «Economic Rationality and Moral Constraints», la concepción de la racionalidad que Gauthier adopta para su teoría moral —la «racionalidad económica»— y el modelo descriptivo de la elección individual basada en tal racionalidad —la Teoría de la Decisión Racional y Teoría de Juegos— habían quedado ya definitivamente fijados. Estaba por añadir, sin embargo, el modelo apropiado para la Decisión Social o Colectiva. En efecto, una teoría de la racionalidad comienza por definir el concepto y describir la acción individualmente racional en todo tipo de contextos, pero ha de concluir con una descripción del comportamiento colectivamente racional o, al menos, con la defensa de un procedimiento de decisión social coherente con la parte de la teoría referida a la racionalidad individual. Pero además, una teoría moral que se concibe a sí misma como una parte, y quizá la más significativa, de la Teoría de la Elección Racional³⁶, sólo puede defenderse si, dada la racionalidad de los individuos y ciertas condiciones fácticas (condiciones de la justicia), existe un procedimiento racionalmente justificado mediante el cual *todos* los individuos sean capaces de decidir un *solo* marco para la interacción cooperativa, al cual denominaremos «justicia» o «moral», siguiendo a Rawls o Gauthier respectivamente. Esto quiere decir que el procedimiento para la elección social, que servirá como modelo ideal para la elección de los principios de la justicia, es uno de los elementos

³⁶ Cfr. John RAWLS, *Teoría de la Justicia*, México, F.C.E., 1979, p. 34.

nucleares de este tipo de teorías³⁷. Como hemos visto, Gauthier se había aproximado al problema de seleccionar el procedimiento ideal para la decisión colectiva al reconocer, al final de «Reason and Maximization», que la elección entre distintos óptimos sociales posibles es una cuestión que sólo se resolvería tras arduas negociaciones entre las partes.

8. EL PÁPEL DE LA «NEGOCIACIÓN RACIONAL»

La idea de negociación será clave, pues Gauthier no acepta el modelo de «decisión colectiva» propuesto por Rawls. De hecho, en su artículo «The Social Contract: Individual Decision or Collective Bargain?»³⁸ («El contrato social: ¿Decisión individual o negociación colectiva?»), de 1978, Gauthier se une a la crítica relativamente común de que en la *Teoría de la Justicia* de Rawls, no se nos presenta un contrato propiamente dicho (por más que el autor se auto-proclame contractualista), sino una decisión racional individual que, porque podría haber sido tomada por (o representa la decisión de) *cualquier* individuo racional en las condiciones de la «posición original», puede tomarse legítimamente como una decisión *unánime*. Pero se trata de una simple decisión, no hay contrato alguno. Gauthier, por el contrario, supone que las condiciones iniciales no serían exactamente las descritas por Rawls³⁹ y, a partir de las que él propone, no habría lugar para una elección individual de principios de justicia. Tales principios habrían de ser negociados partiendo de que son aquellos que regirán la distribución de los bienes sociales (los beneficios de la cooperación) y, como individuos racionales, cada parte pretenderá obtener tanta cantidad de estos bienes como sea posible. Se trata, por tanto, de un problema de negociación.

Afortunadamente, la teoría de la negociación, aún en estado embrionario en los años setenta, había logrado ofrecer una posibilidad de escapar al Teorema de imposibilidad de Arrow, propo-

³⁷ Y uno de sus mayores problemas, pues no olvidemos que cualquier teoría de la Elección Social debe salvar el Teorema de Arrow sobre la imposibilidad de derivar una función de utilidad social a partir de funciones de utilidad individuales.

³⁸ En C.A. Hooker, J.J. Leach y E. F. McClennen (eds.), *Foundations and Applications of Decision Theory*, Boston, Reidel, 1978, vol. II, pp. 47-67.

³⁹ Cfr. arriba nuestra breve explicación del artículo «Justice and Natural Endowment: Toward a Critique of the Rawl's Ideological Framework».

niendo un procedimiento racional para la toma de decisiones en situaciones de intereses contrapuestos. Las contribuciones más importantes a la teoría de la negociación provenían de John F. Nash⁴⁰, Ehud Kalai y Meir Smorodinsky⁴¹, así como del propio Gauthier, quien, en 1974, había desarrollado un procedimiento cuyos resultados coincidían con el de Kalai-Smorodinsky⁴². Así pues, cuando el último peldaño de una Teoría de la Decisión (que era a la vez la dovela clave de una teoría moral contractual) había de ser labrado, Gauthier trajo a escena el concepto de «negociación racional». Éste toma carta plena de naturaleza en su artículo «Social Choice and Distributive Justice»⁴³ («Elección social y justicia distributiva»). Allí discute Gauthier el problema que para la elección social representa el Teorema de Arrow y, basándose en los modelos de negociación de Nash y Kalai-Smorodinsky, propone la negociación como método ideal de decisión social, mostrando que ofrece resultados más acordes con nuestras ideas pre-teóricas sobre la igualdad y la justicia distributiva que aquellos arrojados por el utilitarismo o por los principios de la justicia de Rawls. Su conclusión puede resumirse en las palabras finales del artículo: «Negociar principios aclara el lugar de la igualdad y el carácter de la justicia distributiva, dentro del marco de beneficio mutuo que una sociedad democrática presupone»⁴⁴. El artículo defendía y justificaba, por tanto, el papel (político) de la negociación como medio de decisión social⁴⁵.

Desde ese momento, Gauthier se esfuerza en mostrar el papel de la negociación racional, no ya en una teoría de la decisión social,

⁴⁰ John F. NASH, «The Bargaining Problem», *Econometrica*, 18, 1950, pp. 155-162.

⁴¹ E. KALAI y M. SMORODINSKY. «Other Solutions to Nash's Bargaining Problem». *Econometrica*. 43. 1975. pp. 513-518.

⁴² D. GAUTHIER, «Rational Cooperation», *Noûs*, 8, 1974, pp. 53-65.

⁴³ En *Philosophia*, 7, 1978, pp. 239-253.

⁴⁴ GAUTHIER, «Social Choice and Distributive Justice», cit., p. 252.

⁴⁵ Aunque también reconocía sus límites: la negociación es un procedimiento racional que parte de una situación inicial de negociación que toma como dada. La negociación no ofrece, por sí misma, criterios para juzgar esa situación. De hecho, la negociación sólo es racional entre individuos (o sociedades) que puedan esperar algún beneficio mutuo derivado de una actividad conjunta. Cuando la situación inicial no permite crear esta expectativa, no hay lugar para la negociación sobre la distribución de los bienes sociales ni, por tanto, para la justicia. Debido a esto, Gauthier define como «hobbista» la visión de la justicia que ofrece en este artículo. Posteriormente, la introducción de la cláusula cautelar lockeana como límite normativo de la situación inicial de negociación le permitirá suavizar esta conclusión.

sino en una teoría moral. Dos artículos son fruto inmediato de ese esfuerzo. El primero, «Bargaining our Way into Morality: A Do-It-Yourself Primer»⁴⁶ («Negociar nuestra entrada en la moralidad: manual para principiantes»), es una explicación relativamente asequible de cómo la negociación racional selecciona un resultado óptimo (máximamente beneficioso para cada negociador) de un modo *imparcial*. La negociación es, por tanto, en la teoría de Gauthier, el artificio que permite explicar, casi diseccionándolo, el trayecto que va desde los intereses individuales hasta el pacto social (o, en su caso, el pacto moral). La tradición contractualista ofrece un estado de naturaleza (descrito en mayor o menor detalle) y un pacto hipotético sobre el que los individuos acuerdan. Lo que Gauthier ofrece, gracias a la teoría de la negociación racional, es el proceso mismo de regateo, de exigencias y concesiones mutuas, que conducen a individuos auto-interesados a adoptar un acuerdo mutuamente beneficioso y a someter sus decisiones futuras a límites (morales) que posibiliten el mantenimiento de esa interacción cooperativa beneficiosa basada en el acuerdo.

El segundo artículo sobre la negociación racional es de 1985, y expone ya una versión completa, formalizada y axiomatizada de la teoría, que habría de ser reproducida en el capítulo V de *MA*. «Bargaining and Justice»⁴⁷ («Negociación y justicia») debate no sólo con los modelos de negociación de Nash y Kalai-Smorodinsky, sino también con las axiomatizaciones de Roth⁴⁸, defendiendo la llamada «solución G» (la solución de la negociación basada en la Teoría de la Negociación desarrollada por el propio Gauthier). Por otro lado, el artículo, que representa la etapa de madurez de la teoría moral de Gauthier, sostiene que existe una conexión entre la negociación y la justicia, y proclama, frente a las tesis de Rawls, que la idea de negociación supera ciertas carencias de la «decisión tras el velo de ignorancia» y se ajusta en mayor medida a la definición del individuo liberal.

9. LA INTRODUCCIÓN DEL CONCEPTO DE «ZONA EXENTA DE MORALIDAD»

Por último, debemos mencionar que, una vez completado el mapa de la racionalidad y el nexo de la racionalidad con la mora-

⁴⁶ En *Philosophical Exchange*, 2, 1979, pp. 14-27.

⁴⁷ En *Social Philosophy and Policy*, 2, (1985), pp. 29-47.

⁴⁸ Alvin E. ROTH, *Axiomatic Models of Bargaining*, Berlin, Springer Verlag, 1979.

lidad a través de la negociación, Gauthier introduce, en 1982, una idea que, si bien no juega un papel esencial en su teoría, está incluida en *MA* y aporta claridad y plausibilidad al argumento principal. Se trata de la idea de una «zona exenta de moralidad», identificada con el mercado perfectamente competitivo tal como es concebido por los economistas clásicos. Aparece en el artículo «No Need for Morality: The Case of Competitive Market»⁴⁹ («Moralidad innecesaria: el caso del mercado competitivo»). Este ensayo expone que, dadas las condiciones que los economistas clásicos atribuyen a un mercado perfectamente competitivo, incluyendo la identificación de cada individuo con su función de utilidad y su dotación inicial de factores, así como el «desinterés mutuo», la producción de bienes es óptima, y su atribución a los individuos completamente neutral. Gauthier analiza el resultado de la interacción competitiva (bajo las condiciones ideales de competencia perfecta) y concluye que ningún individuo tendría motivo alguno para quejarse de «parcialidad» en el mercado: todos los individuos son perfectamente libres y el mercado les permite, mediante el intercambio, obtener tanta satisfacción como les cabe esperar, dada su dotación inicial de factores. Si entendemos la moralidad como un mecanismo imparcial para corregir ciertas parcialidades (injusticias) en la distribución de bienes sociales, no jugaría ningún papel en el mercado perfectamente competitivo, pues no habría ninguna parcialidad que corregir.

Desde luego que Gauthier se apresura a reconocer, primero, que las condiciones para la competencia perfecta son, no sólo irrealizables en la práctica, sino rigurosamente imposibles en un mundo tal como lo conocemos; y, segundo, que aunque la operación del mercado sea perfectamente imparcial, la distribución inicial de factores sí está sujeta a valoración moral. Por otro lado, añade que la descripción del mercado ideal como medio perfecto para la distribución imparcial de bienes no implica la defensa de la economía liberal en la práctica. Lo que este artículo, y el capítulo IV de *MA*, quieren poner de manifiesto —más allá de cualquier interpretación simple que los pudiera ver como una defensa de la sociedad de mercado— es la tesis de que la moral y la justicia tienen su lugar propio en el marco de las relaciones sociales reales (y en este marco son necesarias), pero cabe imaginar un contexto de interacción entre individuos racionales en el que los límites morales (en realidad, cualquier límite) serían superfluos, pues la libertad de cada uno conduciría al

⁴⁹ En *Philosophical Exchange*, 3, 1982, pp. 42-55.

mejor estado posible. En definitiva, se trata de un alegato en favor de la prioridad lógica de la racionalidad sobre la moralidad; una hipótesis necesaria para el argumento contractualista, pero que no prejuzga el desarrollo (social, psicológico) ontogenético de la moralidad, ni se refiere a la naturaleza de nuestros sentimientos morales.

El concepto de «zona exenta de moralidad» es el último elemento central de la teoría de Gauthier que podemos rastrear en sus artículos anteriores. En *MA* aparecen, al menos, dos concepciones esenciales más: la idea de una cláusula cautelar que restringe las situaciones iniciales de negociación admisibles entre individuos racionales, y la idea de un punto de Arquímedes, o punto de vista desde el que un individuo cualquiera podría «mover el mundo moral», es decir, diseñar por sí mismo las instituciones y el modo de interacción de un mundo perfectamente imparcial. Estos dos últimos conceptos no se aclararon hasta los últimos estadios de elaboración de la teoría, como el propio Gauthier confiesa en el prefacio de *MA*.

10. EL HALLAZGO DE LA MORAL POR ACUERDO

La revisión de la trayectoria intelectual de Gauthier nos enseña que la teoría moral expuesta en *MA* no es en absoluto un fruto casual, sino el resultado de muchos años de persecución de la posibilidad de fundar una moral basada exclusivamente en los presupuestos más débiles y las concepciones más comunes de nuestra comprensión pre-teórica del mundo práctico. El hecho de que se trate de una teoría contractualista no es más que la consecuencia de una profundización extraordinaria en la comprensión de ciertos aspectos de la racionalidad individual y colectiva y su conexión con la imparcialidad y la justicia. Desde este punto de vista, tanto la moral como su fundación contractual aparecen como necesidades de la racionalidad humana en general.

Precisamente este convencimiento de que el contractualismo es la única vía posible de fundamentación moral, ha conducido a Gauthier a revisar la historia de la filosofía, produciendo interpretaciones originales, no sólo de Hobbes, sino también de Rousseau, Kant o Hume⁵⁰. No nos detendremos aquí en estos aspectos (que podemos llamar colaterales) de su producción bibliográfica;

⁵⁰ Destaca, por lo polémico que resulta, su artículo «David Hume, Contractarian», *Philosophical Review*, 88 (1979), pp. 3-38.

es suficiente con dejar constancia de que, a la vez que nuestro autor ahondaba en el análisis de los métodos y supuestos teóricos esenciales de la moral contractual, también se situaba respecto a sus referentes históricos, dejando claro que el alcance de sus aportaciones no se limitaba a meras críticas coyunturales o técnicas sobre aspectos de la Teoría de la Decisión o la Teoría de la Justicia, sino que formaban parte de un proyecto mayor que pretendía reformular la tradición contractualista que parte de Hobbes.

En esta misma línea, podemos citar un ensayo publicado por Gauthier en 1977, titulado «The social Contract as Ideology»⁵¹ («El contrato social como ideología»), que contrasta con la mayoría de los artículos que hemos venido comentando hasta ahora, pues tiene la forma de un remanso en el caudaloso río de conceptos y explicaciones técnicas con que Gauthier nos inunda durante la década de los setenta. En este ensayo aparecen sus intuiciones generales sobre el individualismo y la racionalidad propios del modelo contractualista, pero hay sobre todo una reflexión detenida sobre las implicaciones mutuas entre el pensamiento (o ideología) contractualista y la sociedad occidental de mercado en la que vivimos. En esa reflexión, Gauthier encuentra que el individualismo posesivo, la racionalidad maximizadora o la concepción de las relaciones sociales como contractuales yacen en el centro mismo de nuestra ideología sobre la sociedad y el hombre, y que no son desligables una de otra. De este modo, percibe el contractualismo como el modelo de fundamentación de las instituciones sociales al que estamos abocados, dada nuestra visión del mundo. Pero, al mismo tiempo, detecta que la estabilidad efectiva de las instituciones sociales se basa, no en el interés particular que las justifica teóricamente, sino en sentimientos aparentemente irracionales desde el punto de vista de nuestra propia ideología (el amor, el patriotismo, etc.).

Por eso, el conjunto de los avances técnicos de la Teoría de la Decisión Racional, la influencia de las tesis de Baier, el lejano magisterio de Hare, con su empeño de racionalizar la moral, la fascinación por el *Leviathan* de Hobbes y el candente debate con el neocontractualismo de Rawls y Nozick, abocaron a Gauthier a intentar lanzar el puente entre los elementos indiscutibles de nuestra ideología y nuestras convicciones como sujetos morales.

⁵¹ En *Philosophy and Public Affairs*, 6, otoño 1977, pp. 130-164.

11. GAUTHIER EN LA ÉTICA CONTEMPORÁNEA

A lo largo de su empresa —en los años que van desde 1969 hasta 1986— Gauthier se instala en la corriente principal del pensamiento neo-contractualista, asumiendo y desarrollando conceptos y métodos difundidos por Rawls, Buchanan, etc., pero contando, como aportación distintiva, con una sólida formación analítica y una gran fidelidad a los supuestos filosóficos del contractualismo de Thomas Hobbes. Desde estos orígenes, y a través del proceso que hemos ido siguiendo, Gauthier encuentra en el contractualismo el medio para afrontar el problema que le inquietaba desde sus inicios: la posibilidad de proporcionar un fundamento racional para la moral. En *Morals by Agreement* se defiende la tesis de que una explicación contractualista de la moral puede articular las exigencias de la racionalidad como maximización y las demandas de la justicia, así como asegurar un criterio definitivo de corrección moral.

Morals by Agreement se recibió en los países anglosajones como un capítulo más de la productiva serie neo-contractualista iniciada por la revolucionaria (en su momento) *Teoría de la Justicia* de J. Rawls. Pronto se advirtió lo ambicioso de su intento, que no se limitaba, como en el caso de otros contractualistas, a propugnar una fundamentación consensual del Estado o de la Justicia en el marco de la Sociedad, sino que pretendía justificar filosóficamente la racionalidad de la moral, mediante el recurso teórico del contrato. El interés que despertó la obra queda reflejado en los números especiales que le dedicaron prestigiosas revistas y en una serie de simposios y seminarios sobre la misma⁵².

En España, la obra de Gauthier fue recibida en los ámbitos iusfilosófico y ético familiarizados con sus antecedentes neo-contractualistas y también con ciertos aspectos de la teoría de la decisión racional⁵³. Predominó, entre los pocos especialistas que se acerca-

⁵² Entre las revistas que han editado números especiales destacan el *Canadian Journal of Philosophy*, que le dedicó su número 18, y *Ethics*, que en su número 97 recoge un *simposium* sobre la obra de Gauthier. Otros congresos dieron lugar a monografías, como E.F. Paul *et al.* (eds.), *The New Social Contract: essays on Gauthier* (Oxford, Blackwell, 1998); P. Vallentyne (ed.), *Contractarianism and Rational Choice: Essays on David Gauthier's Morals by Agreement*, (New York, Cambridge U.P., 1990); D. Gauthier y R. Sugden (eds.), *Rationality, Justice and the Social Contract. Themes Morals Morals by Agreement* (Ann Arbor, University of Michigan Press, 1993).

⁵³ Los comentarios más relevantes a *Morals by Agreement* en español se encuentran en *Doxa*, n.º 6 (1989), que dedica a la obra de Gauthier la parte

ron a esta obra, una doble valoración: por un lado se ponderaba el esfuerzo del autor, por otro, se reputaba inalcanzado su objetivo. La mayoría de los comentaristas coincidían en señalar los problemas que Gauthier habría de enfrentar para demostrar que es «racional ser moral» sin transformar ilegítimamente alguna de sus premisas, es decir, sin traicionarse a sí mismo. En general, los comentarios españoles detectan los aspectos más problemáticos de la teoría de Gauthier, pero sostengo que desechan demasiado pronto su proyecto, sin advertir la potencia filosófica del mismo. Achaco esta carencia al hecho de que no ha habido en España una lectura verdaderamente filosófica de la obra de Gauthier. Creo que si esa lectura se lleva a cabo, se detectarán, sin duda, los problemas ya señalados, pero se destacará también, de modo más vehemente, el esfuerzo que representa este intento. A nuestro juicio es una de las formulaciones meta-éticas contemporáneas de más hondo calado y de mayor proyección futura; pero esto es, por el momento, sólo una opinión. Esperamos que de esta breve genealogía de la moral por acuerdo puedan extraerse algunas razones a favor de esa opinión. En todo caso, confiamos en que haya mostrado el interés filosófico de la teoría ética más ambiciosa del momento: aquella que promete fundar una moral universal sobre la escueta base de la mera razón instrumental.

monográfica del número, recogiendo artículos de R. Zimmerling, M. D. Farrell y A. Calsamiglia, y en la *Revista de Filosofía*, 3.^a época, vol. IV (1991), n.º 5, que añade una interesante recensión de J. Montoya; asimismo, destaca el tratamiento de J.C. Bayón Mohino en su obra *La normatividad del derecho: Deber jurídico y razones para la acción* (Madrid, Centro de Estudios Constitucionales, 1991) y el análisis de Martín D. Farrell en *La filosofía del liberalismo* (Madrid, Centro de Estudios Constitucionales, 1992, especialmente pp. 79-86). En 1992, Dña. M^a Pilar González Altable publica, en esta misma revista, «El contractualismo liberal de D. Gauthier: Contractualismo vs. Utilitarismo» (*Τέλος*, vol I, n.º 2, Junio 1992, pp. 111-127). González Altable se hace eco, en parte, de una de las líneas de investigación del Dpto. de Filosofía del Derecho Moral y Política II (Ética y Sociología) de la U.C.M., centrada en la Teoría económica de la Decisión Racional y Juegos como desarrollos post-analíticos relevantes para la ética utilitarista y contractualista. En el marco de esta línea de investigación, la ética de Gauthier había sido ya tratada en la Tesis Doctoral de Dña. Blanca Rodríguez López, *Moralidad y cooperación racional*, dirigida por el Prof. Gilberto Gutiérrez y leída en 1991. Posteriormente, tanto la Dra. Rodríguez como yo mismo, hemos ofrecido diversos seminarios sobre la obra de Gauthier: «Racionalidad, Moralidad y Contrato Social en la Obra de David Gauthier» (curso 1992-93); «Lectura de la *Moral por acuerdo* de David Gauthier» (curso 1993-4). En ellos se ha discutido la posibilidad de una lectura propiamente filosófica del intento de Gauthier, como alternativa al sesgo jurídico que se observa en la mayoría de los artículos publicados.