UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE FILOSOFÍA

DEPARTAMENTO DE FILOSOFÍA DEL DERECHO, MORAL Y POLÍTICA II

TESIS DOCTORAL

El Contrato Moral: Individuo, racionalidad, acuerdo y la Teoría Etica de David Gauthier

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Pedro Frances Gómez

DIRIGIDA POR

Gilberto Gutiérrez López

Madrid, 2002

ISBN: 978-84-8466-222-8

FACULTAD DE FILOSOFÍA
UNIVERSIDAD COMPLUTENSE DE MADRID
Departamento de Filosofía del Derecho, Moral y Política II (Ética y Sociología)

El Contrato Moral

Individuo, racionalidad, acuerdo,
y la teoría ética de David Gauthier

TESIS DOCTORAL QUE PRESENTA D. PEDRO FRANCÉS GÓMEZ,
BAJO LA DIRECCIÓN DEL DR. D. GILBERTO GUTIÉRREZ LÓPEZ,
CATEDRÁTICO DE ÉTICA Y DIRECTOR DEL DEPARTAMENTO DE
FILOSOFÍA DEL DERECHO, MORAL Y POLÍTICA II (ÉTICA Y
SOCIOLOGÍA) DE LA UNIVERSIDAD COMPLUTENSE DE MADRID.

"Edifiquemos con palabras una ciudad desde sus cimientos. La construirán, por lo visto, nuestras necesidades."

Platón, *La República*, 369c.

Índice

Agradecimientos 1

Introducción

1.- ¿Por qué contractualismo? 5
2.- ¿Por qué David Gauthier? 8
3.- Planteamiento del trabajo 9

Capítulo I:
Presentación: David Gauthier y el desarrollo de la moral por acuerdo 13

1.- Las preocupaciones iniciales y la influencia de Hare 14
2.- La primera obra: prudencia, moralidad y lenguaje 16
3.- Hobbes y la Teoría de Juegos: el punto de partida 19
4.- El impacto de la *Teoría de la justicia* de Rawls 22
5.- El concepto de "maximización restringida" 27
6.- Racionalidad económica y moralidad 30
7.- El papel de la "negociación racional" 37
8.- La introducción del concepto de "zona exenta de moralidad" 40
9.- El hallazgo de la moral por acuerdo 42
10.- Gauthier en la ética contemporánea 44

Capítulo II:
Presupuestos del contractualismo moral 49

1.- El postulado del individualismo 56
a) El origen del individualismo 58
b) Individuo y contrato 64
c) El individualismo metodológico 69
d) La versión de Gauthier: primeras ideas 75
e) Individuos mutuamente desinteresados 81
f) Individuo económico y "yo de mercado" 94
g) Excurso: el referente rawlsiano 106

2.- La racionalidad de las partes 121
a) Advertencias iniciales 121
b) Racionalidad instrumental, consecuencialismo y prudencia 131
c) Racionalidad y maximización 141
d) Principios de decisión racional 151

e) La hipótesis de igual racionalidad 155
f) Excurso: paradigmas de racionalidad 160
g) La moral como parte de la Teoría de la Decisión Racional 170

3.- Preconcepción de la moral y teoría del valor 175

4.- El papel del mercado 188

5.- Premisas teóricas y estado de naturaleza 207

Capítulo III:
Hipótesis histórica: ¿Hay una tradición de contractualismo moral? 219

1.- Presentación de nuestra hipótesis 219

2.- Diferencia entre convencionalismo y contractualismo 227
a) La confusión de los términos 227
b) Puntos de partida comunes 232
c) Bases para la distinción 236
d) Distinciones complementarias 243
e) Motivación, conocimiento y explicación 246

3.- Contractualismo político y convencionalismo moral 253
a) Glaucón, ¿el primer contractualista? 253
b) *Communitas* vs. *societas*: el debate medieval 257
c) La revolución hobbesiana y sus consecuencias 261
d) La ilustración: Hume, Rousseau y Kant 267
e) Lectura de la tradición del contrato social 274

4.- Interpretación del contractualismo de Hobbes 276
a) Posibilidad de un contractualismo moral en Hobbes 276
b) Resumen de nuestra interpretación 278
c) El punto de partida del argumento 283
d) La solución del conflicto natural 292
e) Razón común y moralidad 295
f) Excurso: la ley natural o la ética de Hobbes 302
g) El camino hacia el contractualismo moral 306

5.- Sobre los antecedentes del contractualismo moral liberal 312

Capítulo IV:
El contrato moral 321

1.- El contenido del contractualismo moral 321
a) Justificación moral y contrato 322
b) El contrato moral y las teorías liberales de la justicia 327
c) Estructura de la teoría del contrato moral 331

2.- La interacción natural y el "Dilema del Prisionero" 337
a) La irracionalidad de elegir el egoísmo 337
b) La racionalidad de la cooperación 343
c) Excurso: estado de naturaleza y coordinación 347

3.- Negociación racional y pacto 351
a) Optimización, negociación y contrato social 352
b) El objeto de la negociación 354
c) Sobre el procedimiento de negociación 357
d) El principio de concesión relativa *minimax* 366
e) La utilidad relativa *maximin* 372
f) Devenir y crítica del principio 374

4.- La racionalidad como maximización restringida 386
a) ¿Qué exige la cooperación? 392
b) El argumento en favor de la maximización restringida 397
c) Las objeciones de los herederos del *Tonto* 404
d) Transparencia y translucidez 413
e) Qué clase de disposición es racional adoptar 419
f) ¿Disposiciones, resoluciones o planes? 426
g) La racionalidad de cumplir los pactos 431

5.- Contrato social y derechos individuales 436
a) El papel de la salvaguardia 438
b) Libertad y derechos en el estado de naturaleza 447
c) Las dimensiones de la salvaguardia 452
d) Conclusión: la salvaguardia y la cooperación 455

6.- Moral contractual e individuo: recapitulación final 459

Capítulo V:
Conclusión: Razón pública y moralidad 469

1.- La moral de una sociedad liberal 469
2.- Contrato moral, universalismo y racionalidad 472
3.- Ética mínima, liberalismo y el ámbito de lo público 475
4.- El contrato moral como ideología: crítica final 478

Apéndice:
Breve revista de la recepción de la teoría de David Gauthier en la filosofía española 485

Bibliografía 499

Agradecimientos

La presente tesis doctoral ha sido realizada con el apoyo financiero e institucional de la Universidad Complutense de Madrid —a través de una beca predoctoral (Programa de Formación de Personal Investigador y Docente)— y del Departamento de Filosofía de Derecho, Moral y Política II (Ética y Sociología) de la misma Universidad. Debo mencionar especialmente al Prof. Gilberto Gutiérrez López quien, en su calidad de Director del Depto., ha puesto a mi disposición todos los medios existentes para favorecer mi investigación.

Ni con mucho es ésta la única deuda que he contraído con el Prof. Gutiérrez, director de esta tesis. Él me introdujo en un método de la filosofía práctica, no sólo prometedor e inestimable como herramienta para la reflexión, sino además sorprendentemente acorde con mi propio enfoque de la filosofía, y aun con mi talante y disposición personales. Después, me condujo hacia la formulación definitiva de un problema que ni siquiera sospechaba cuando me interesé por el contrato social como materia de estudio. Más tarde, ha revisado pacientemente los borradores de mi trabajo, ayudando a mejorarlos con sus comentarios y correcciones. Tan agradecido le estoy por todo ello como por el ánimo constante y el apoyo que he tenido siempre de su parte.

Agradezco especialmente al Prof. David Gauthier su amable acogida y sus comentarios que, además de mejorar mi comprensión de muchos temas, han orientado la formulación definitiva de mi trabajo. Su aprobación de gran parte de mis ideas, incluso las que eran críticas con su proyecto, fue de gran importancia para mí y me impulsó a aplicarme en la redacción del texto final.

También quiero dar las gracias al Prof. Carlos Thiebaut por sus consejos, su ánimo y sus valiosísimos comentarios a algunas partes del manuscrito; a los profesores Celia Amorós y J. Miguel Palacios por su magisterio y su ejemplo; a Angeles Jiménez Perona, Mercedes Gómez Adanero y José Luis Muñoz de Baena por sus sugerencias y comentarios durante la primera fase de mi trabajo; a Blanca Rodríguez López, por señalarme el camino y compartir conmigo inquietudes, tiempo y material bibliográfico; a los profesores Albert Calsamiglia y Manuel Atienza por atender mis consultas al comienzo de la investigación; a mis compañeros Teresa Padilla, Javier de la Torre, Alejandro Escudero, Agustín Lejarreta, J.A. Valor y Javier Alonso por sus opiniones y por muchas otras cosas; y a los alumnos de cuarto y quinto que han soportado mis seminarios en la Facultad, por sus fructíferas críticas.

Mención aparte merecen los desvelos de Julia García Maza, directora de la Biblioteca del Instituto de Filosofía del CSIC, así como el equipo que dirige, por asegurar que mi información bibliográfica y documental se encontraba —y se encuentra— permanentemente al día. Mi agradecimiento es tanto como su celo, que sobrepasa con mucho lo profesional.

A todas estas personas, a quienes he transferido las muchas "externalidades negativas" que produce este trabajo, y a otras que sin duda olvido, dedico el pequeño e imperfecto fruto de mi esfuerzo.

Introducción

Introducción

1.- ¿Por qué contractualismo?

> "La moralidad se ve enfrentada con una crisis de fundamentación. El contractualismo ofrece la única solución plausible de esta crisis"[1].

Ésa es la tesis que nos proponemos defender. Y la pregunta es inmediata: concedamos que la moralidad está enfrentada a una crisis de fundamentación, ¿por qué es el contractualismo la única solución a esa crisis? Para contestar, deberíamos aclarar primero qué entendemos por contractualismo; definido éste, podremos justificar su plausibilidad como "solución" a la crisis de la moralidad.

El contractualismo es, ante todo, un método. Un método de investigación propio de la filosofía práctica que nos permite distinguir con claridad las demandas de la razón en su uso moral, es decir, en todo lo que concierne a la interacción (real o posible) con otros agentes. El 'contrato social' es un

[1] Gauthier, D., "¿Por qué contractualismo?", *Doxa*, 6 (1989), pp. 19-38; p. 19.

argumento complejo que permite justificar racionalmente, ante el individuo, ciertas restricciones en su comportamiento (tales como la obediencia al poder político, el sometimiento a ciertas instituciones sociales o el respeto a las reglas y principios morales). De este modo, el contractualismo ayuda a que el individuo no vea como arbitrarias estas restricciones, sino que las tenga por requisitos de *su* racionalidad, en cuanto relacionada con las expectativas y acciones de otras personas.

El argumento contractualista suele desarrollarse en tres fases: La primera es la definición de un escenario hipotético que reúne ciertas condiciones plausibles referidas al carácter de los individuos que lo pueblan, su racionalidad y su relación con el medio; la segunda consiste en analizar la interacción en el "estado de naturaleza" y sus posibles soluciones; por último, una tercera fase emplearía las conclusiones de la segunda para defender el estátus normativo de la autoridad política o moral.

Un argumento contractualista será válido si puede derivarse lógicamente de las premisas; además, será plausible si las premisas mismas lo son[2].

Como tal método, el contractualismo no pre-determina el carácter de su conclusión normativa. De hecho, la plausibilidad del argumento dependerá de definir unas premisas cuyo componente normativo tienda a cero, y tratar de deducir conclusiones normativas siguiendo un procedimiento lógicamente impecable. Si ambas condiciones se cumplen, el resultado del contractualismo podría incluso contrastar fuertemente con nuestras convicciones morales. Nuestro interés no se centrará, sin embargo, en las eventuales conclusiones del argumento, sino en su posibilidad misma, como método de la filosofía moral.

Quizá sea útil ofrecer un ejemplo del posible funcionamiento del contractualismo como filosofía moral. Thomas M. Scanlon nos ofrece el siguiente, referido a la explicación contractualista de la incorrección moral[3]:

---

[2] Respecto a este esquema del contractualismo, así como a las categorías de validez y plausibilidad, cfr. Kraus, J., *The Limits of Hobbesian Contractarianism*, Nueva York, Cambridge U.P., 1993, *Introducción*.

[3] El ejemplo se encuentra en "Contractualism and Utilitarianism" (en Sen y Williams (eds.), *Utilitarianism and Beyond*, Nueva York, Cambridge U.P., 1982, pp. 103-128), p. 110, y alude a una concepción del contractualismo un tanto diferente de la que nosotros defenderemos; pero

"Un acto es incorrecto si su realización en las circunstancias dadas no estaría permitida por un sistema de normas de regulación general de la conducta que nadie pudiera rechazar razonablemente como base para un acuerdo general informado y no coactivo". De este ejemplo se deduce que una teoría moral contractual debe especificar las condiciones y contenido de un acuerdo general razonable. El objetivo último del contractualismo, al cual se encaminaría el argumento que hemos esquematizado arriba, es (en este caso) proporcionar al individuo un criterio normativo de corrección moral.

Estas notas incompletas tal vez ocasionen más confusión que claridad, pero bastarán por el momento para avanzar nuestra concepción del contractualismo y su papel como filosofía moral. Ahora podemos pasar a la cuestión relevante, ¿por qué nos parece que éste método es el único viable para solucionar la crisis de fundamentación de la moralidad?

La razón básica es el compromiso estricto del contractualismo con el empleo de premisas mínimas, incontrovertibles, convincentes para la mayoría de las personas. Estimamos que sólo un argumento basado en tales premisas puede esperar ser aceptado (y generalmente aceptable) en el marco de una sociedad liberal, democrática y de mercado. Desde nuestro punto de vista, el contractualismo no sólo es un método de la filosofía moral, sino el único modo de expresión de la razón práctica en un marco esencialmente plural. A lo largo del trabajo, intentaremos dar argumentos en favor de esta visión.

---

puede servir como orientación inicial.

2.- *¿Por qué David Gauthier?*

Varios motivos nos han llevado a centrarnos en el contractualismo liberal de David Gauthier. No es el menor de ellos el hecho de que representa la última contribución relevante a la escuela "neo-contractualista", encarnada en autores como Rawls, Buchanan, Nozick, Scanlon y otros. Pero la razón fundamental es que Gauthier eleva el argumento contractualista a la más alta cota de abstracción y refinamiento[4]. De modo que, si nuestra tesis no es defendible sobre la base del argumento de Gauthier, podríamos renunciar completamente a ella. La propuesta de Gauthier puede considerarse como "la prueba definitiva" del contractualismo en general; si el contractualismo no es plausible en esta forma ¿de qué otra forma podría serlo?

Por otro lado —y en íntima conexión con el hecho de tratarse de la versión más sofisticada del contrato social— argumentaremos que la teoría de Gauthier supone uno de los escasos de ejemplos satisfactorios de contractualismo moral. En realidad, es prácticamente el único intento explícito de ofrecer una justificación contractual de la moralidad.

Nuestro intento es hallar y explorar el camino de salida de la crisis de fundamentación moral. Al elegir el contractualismo liberal lo único que hacemos es seguir la senda que nos parece más prometedora —y en esa travesía, estábamos abocados a fijarnos en el modelo de Gauthier. Nuestro trabajo consistirá, por tanto, en analizar, punto por punto, los elementos de esta teoría, para intentar averiguar si efectivamente nos conducen hacia donde prometen.

---

[4] Al decir de David Braybrooke, "la teoría del contrato social vuela, en *Morals by Agreement*, más alto y con más pericia que nunca. Pero no vuela sola. Gauthier ha llevado a la misma altura de sofisticación el proyecto, a menudo cuestionado entre filósofos de todas las épocas, de deducir la moralidad a partir de la racionalidad. La teoría del contrato social es el vehículo más prometedor para ese proyecto de deducción, al reunir un conjunto de temas cruciales todos ellos para la teoría ética, como el consentimiento, el beneficio mutuo y la cooperación. Nadie se había acercado tanto como Gauthier a llevar a cabo el proyecto con perfecto rigor y precisión." ("Contract Theory's Fanciest Flight", en *Ethics*, 97 (julio 1987), pp. 750-764;p. 751).

3.- Planteamiento del trabajo

El objetivo del trabajo se logrará presentando la filosofía moral de David P. Gauthier en su contexto teórico e histórico, para intentar una interpretación genuinamente filosófico-práctica de la misma; es decir, una interpretación realizada desde el punto de vista de la reflexión ética[5].

La filosofía de David Gauthier no es desconocida en España e hispanoamérica, donde ha alcanzado cierto eco entre filósofos del Derecho y la Política, en parte por su adscripción al neocontractualismo, que es tal vez, en sus distintas versiones, la corriente de pensamiento político más influyente desde finales de los años setenta. Sin embargo, entre los especialistas en Ética, la repercusión de la obra de Gauthier ha sido menor hasta el momento (aunque es apreciable un cambio progresivo de actitud[6]). Ello puede deberse al lenguaje empleado por Gauthier, un lenguaje que debe más a la ciencia económica y social que a la ilustre tradición hobbesiana en la que su teoría se inspira. Además, la obra principal de Gauthier, *Morals by Agreement* (*Moral por acuerdo*), es extensa y compleja. Contiene ante todo una teoría contractualista de la moral, pero también defiende una teoría subjetivista del valor; justifica una concepción concreta de la racionalidad, que luego reformula con reconocida originalidad; parte de una visión del individuo y la persona moral enraizada en la tradición liberal, pero guarnecida con nuevas posibilidades; emplea métodos formales procedentes de la Teoría de la Decisión Racional y de Juegos y, finalmente, propone las líneas generales para una acción política concreta plenamente diferenciada. Esta complejidad e interdisciplinariedad han contribuido a la proliferación de lecturas muy dispares, que alejan lamentable-

---

[5] Defenderemos esta interpretación frente a la mayoría de visiones (casi todas ellas parciales) de la obra de Gauthier ofrecidas hasta el momento.

[6] Como prueban, por ejemplo, los trabajos de Rubio Carracedo ("Los dos paradigmas de la ética: estrategia y comunicación", en Rubio Carracedo, J, *Ética constructiva y autonomía personal*, Madrid, Tecnos, 1992) y Carlos Thiebaut (*Los límites de la comunidad*, Madrid, Centro de Estudios Constitucionales, 1992), los cuales inician lo que podría ser un tratamiento sin prejuicios de la obra de Gauthier en el marco de la ética contemporánea. Por otro lado, la reciente traducción de *Moral por acuerdo* (Barcelona, Gedisa, 1994) supondrá sin duda un impulso en este sentido.

mente la obra de Gauthier de una consideración específicamente ética y filosófica. Nuestro propósito es colaborar en la superación de esta carencia.

El mayor obstáculo en la tarea de justificar y debatir el interés de la obra de Gauthier para la Ética es que tanto su figura como su lenguaje filosófico, aunque conocidos, son poco familiares para quienes reflexionan habitualmente desde otras tradiciones. Nuestro primer objetivo será, por tanto, introducir la figura intelectual de David Gauthier, y —en la medida de lo posible— su lenguaje filosófico y las intuiciones fundamentales de su teoría. A lo largo de ese primer capítulo de presentación completaremos, además, la justificación del interés filosófico del contractualismo liberal.

Tras esta presentación, el capítulo segundo está dedicado a los elementos básicos de la teoría del contrato social (sus premisas): individualismo, racionalidad, concepción del valor, etc. Con ello se inicial el despliegue de la teoría, aunque aún los rasgos específicos del contractualismo moral no aparecerán con claridad. Éstos se apreciarán sólo al exponer detalladamente el núcleo argumental, consistente en el análisis de la interacción natural y el paso a la sociedad (y a la moralidad, en nuestro caso) a través de la idea de una negociación racional. Este análisis ocupará el capítulo cuarto; así, los capítulos segundo y cuarto, junto con el capitulo de conclusiones, forman la línea argumental sistemática de la tesis. El capítulo tercero es un interludio histórico, en el que ponemos en relación el contractualismo moral con sus posibles antecedentes, y defendemos explícitamente lo que en el resto del trabajo (incluida esta introducción) se supone: que la teoría de Gauthier representa quizá la única forma histórica de contractualismo moral propiamente dicho.

Capítulo I

Presentación: David Gauthier y el desarrollo de la Moral por acuerdo

Estas páginas iniciales han de servir como presentación de la figura del Profesor David P. Gauthier, así como de su teoría moral. Creemos que el mejor modo de hacer esta presentación será narrar brevemente su trayectoria intelectual, lo que nos permitirá ofrecer una primera justificación del interés filosófico de su obra y exponer, de paso, las razones que nos animan a llevar a cabo su estudio. En esta presentación introduciremos inevitablemente conceptos cuyo sentido e importancia teórica sólo podrán ser valorados más adelante, una vez precisada su filiación y su función en la teoría moral de Gauthier. De momento sólo los aludiremos, al referirnos a lo que podemos llamar su "nacimiento histórico". Pedimos condescendencia para estos párrafos iniciales, en los que se dará entrada a términos procedentes de la Economía o la Teoría de Juegos sin aclarar completamente su uso filosófico. Esperamos que éste quede patente cuando expliquemos la base teórica de Gauthier y formulemos sistemáticamente su teoría.

1.- 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.

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*[1] 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-cognoscitivismo, 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

[1] 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.

un intento de conectar dos discursos que, desde la perspectiva analítica, parecían irremediablemente alejados.

La necesidad de esta conexión había sido ya reconocida por Hare[2] frente a las entonces influyentes tesis del existencialismo, el emotivismo y otras formas de no-cognoscitivismo 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 lógico[3]. La universalizabilidad de los juicios morales garantizaba, según Hare, su racionalidad, pues los sometía a un principio lógico (de universalización) que él consideraba esencialmente 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.

---

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

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

2.- 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*[4] (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 "racionalidad" 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:

---

[4] "Razonamiento práctico: estructura y fundamentos de los argumentos prudencial y moral y su ejemplificación en el discurso".

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. 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 la 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 particulares[5]. Tal concepción de la inferencia práctica le permitió 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 aquél primer ensayo

[5] Cfr. *Practial Reasoning*, p. 169.

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 práctico[6]. 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 académico[7]), 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.

---

[6] 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 esa: 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.

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

3.- Hobbes 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 moral[8]. 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*[9], obra que pone de manifiesto la fascinación que el filósofo inglés ejerce sobre Gauthier (Hobbes será un tema recurrente en su bibliografía) y re-descubre su filosofía política y *moral*, lo cual habría de tener repercusiones posteriores[10]. 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)[11]. 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

---

[8] 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.

[9] "La lógica de Leviathan: la teoría moral y política de Thomas Hobbes".

[10] 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, *Hobbesian Moral and Political Theory* (Princeton, NJ, Princeton University Press). Ambas obras deben mucho a la rehabilitación de la filosofía moral hobbesiana emprendida por Gauthier.

[11] Este 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.

de Gauthier en que emplea el Dilema del Prisionero[12], así como rudimentos de Teoría de Juegos, 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 1965 en que Howard Sobel le presentó el Dilema del Prisionero y algunas de las ideas básicas de la Teoría de Juegos[13]. Gauthier estaba ocupado, como ya hemos señalado, en encontrar un modo conveniente de formular la relación entre la moralidad y el interés, a partir de la concepción de moralidad de Kurt Baier. 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[14].

---

[12] Para una introducción rápida, sencilla y sugerente del Dilema del Prisionero y su relevancia para la teoría de la racionalidad y la filosofía moral, puede verse el breve artículo de Derek Parfit "Prudencia, moralidad y el dilema del prisionero", en *Diálogo filosófico*, 13 (1989), pp. 4-30.

[13] 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.

[14] Merece la pena transcribir cómo recuerda Gauthier aquél momento: "Cuando consideré por primera vez la concepción de la moralidad de Kurt Baier, me encontré tratando de comprender el conflicto entre las razones de auto-interés y las razones superiores, y escribí, y leí, una comunicación en la cual los problemas se perdían en un laberinto de palabras. Tras escuchar aquellas palabras, Howard Sobel me llevó aparte y, dibujando una matriz en una hoja de papel, me dijo, "¡Mira!, estás hablando del Dilema del Prisionero". Yo miré y fue como si un velo cayera de mis ojos y recibiera la luz. Este incidente ocurrió en la Universidad de California, Los Angeles, en Noviembre de 1965." (Gauthier, D., "The Incompleat Egoist", en McMurrin S.L. (ed.) *The Tanner Lectures on Human Values*, vol. 5, Salt Lake City, University of Utah Press, 1984, pp. 67-119, reimpreso en Gauthier, D. *Moral Dealing*, Nueva York, Cornell U.P., 1989, pp. 234-273,

En el artículo "La moral y la ventaja" distinguimos ya una formulación bastante 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 moralidad sólo era entonces, para Gauthier, un problema: la solución teórica ni siquiera era aún vislumbrada, ya que aquél 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 aquél 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.

---

por donde se cita: p. 254-255).

4.- 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 racional[15], 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"[16] ("Justicia 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 Rawls[17], conducía, no a los conocidos principios defendido en

---

[15] Esta tesis está recogida en su antológico artículo "The Impossibility of Rational Egoism", *Journal of Philosophy*, 71 (1974) 439-456.

[16] 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.

[17] 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.

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 retribución 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 destinados a jugar un papel central en 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.

El motivo por el que Gauthier no acepta el principio de la diferencia en su versión rawlsiana es que cree que los individuos que lo acordaron en la posición original, una vez levantado el velo de ignorancia, no seguirían viéndolo como un principio de justicia racional. 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 diferencia 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[18]. 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, excluyendo 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

[18] Esto significa variar la *baseline* o posición original respecto a la cual se mide el beneficio que la sociedad representa para cada individuo.

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 el 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*[19], 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 cantidad de bienes primarios, sino de la proporción en que cada individuo ve cumplida su expectativa máxima. Así pues, el

[19] 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.

"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 proclama, 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"[20].

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. Sería necesario que una nueva formulación de la racionalidad, capaz de superar los estrechos límites de la concepción maximizadora, entrara en escena y resultara tan plausible como aquella. Gauthier intentará, precisamente, aportar esa nueva formulación en sus siguientes ensayos, pero antes de exponer ese intento, creo conveniente resumir el significado global de la crítica de Gauthier a Rawls que acabamos de analizar.

Como hemos visto, Gauthier se enfrenta al contractualismo de Rawls sin renunciar al paradigma contractualista, pero afirmando de un modo más radical

[20] "Justice and Natural Endowment...", p. 169.

el origen liberal del mismo. Incluso se percibe cierto legado lockeano[21] en conceptos tales como dotación 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 comentamos. Lo más interesante es que en este ensayo Gauthier da forma a una versión del contrato muy diferente de la rawlsiana, 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.

5.- El concepto de "maximización 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")[22], y estaba destinada a jugar un papel esencial en la teoría moral posterior de David Gauthier. De hecho, es probablemente la concepción clave, y una de las mayores originalidades, de su obra. Habremos de ocuparnos con detalle de su versión definitiva en *MA*, así como de las muchas y muy solventes críticas que se han dirigido contra ella. Pero también será ilustrativo destacar ahora algunos rasgos de su primera formulación.

---

[21] Aunque en modo alguno debe confundirse la teoría de Gauthier con una teoría retributiva de la justicia, del tipo de la de Nozick. Más adelante discutiremos este punto.

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

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*[23]. 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 ejemplo[24] 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

[23] Así aparece enunciado en Gauthier, "Reason and Maximization", p. 216.

[24] 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, aun tomando pie en la realidad e intuitivamente ilustrativo, es una situación ficticia que ha de cumplir perfectamente ciertas condiciones hipotéticas.

cantidad de aquello que el agente prefiera, o la misma cantidad durante más tiempo, 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, ésta segunda acción se adecua 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 Dilema 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 persona con la que interactúa una utilidad tal que no hay ninguna combinación de acciones posibles, una para cada persona de las que interactúan, con un resultado esperado que proporcione a cada una de las otras 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"[25].

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 es (pues un maximizador directo elegiría, guiado por el criterio de su principio racional, un tipo de racionalidad diferente).

[25] "Reason and Maximization", Cit., p. 277-278.

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 (básicamente coincidente con el concepto de justicia de Rawls). Con ello, la moralidad (al menos así entendida) entraría a formar parte de las 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"[26].

6.- 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". De hecho, como tendremos oportunidad de comprobar, el concepto de maximización restringida aparece con muy pocas variaciones en *MA*. 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, publicados casi contemporáneamente.

[26] *Ibid.*, p. 232-233.

El primero de ellos es "Rational Cooperation" ("Cooperación racional")[27], en el que se había intentado satisfacer la demanda (con que acaba "Reason and Maximization") de un procedimiento racional de negociación para alcanzar acuerdos sobre los conjuntos de estrategias que podrían conducir a resultados sociales óptimos. "Rational Cooperation" complementa, así, el punto de "Reason and Maximization" en que, reconociendo que es más beneficioso en términos directamente maximizadores adoptar una estrategia común optimizadora, hay que elegir conjuntamente cuál de las varias estrategias optimizadoras posibles se pone de hecho en práctica. El procedimiento para tal elección común sólo puede consistir en una negociación racional[28], único modelo de toma de decisiones colectivas que logra a la vez escapar al Teorema de Arrow[29] y tomar en cuenta los intereses individuales de todos y cada uno de los agentes dispuestos a cooperar. Más abajo nos referiremos a la importancia de la negociación racional para el contractualismo moral liberal de Gauthier. De momento diremos únicamente que "Rational Cooperation" intenta explicar qué principio regiría la cooperación racional y, al hacerlo, anticipa un principio orientador de la negociación (pues la negociación es el medio que conduce a una estrategia conjunta cooperativa). Tomando como fuente las teorías de la negociación de Harsanyi, Luce y Raiffa y, especialmente, las de Nash y Zeuthen[30], Gauthier desarrolla su propio principio de cooperación racional y contribuye al desarrollo de la, por entonces embrionaria, teoría de la negocia-

---

[27] En *Noûs*, 8 (1974), pp. 53-65.

[28] Es importante apuntar que dicha negociación racional ha de ser planteada en los términos de la racionalidad directamente maximizadora. Aunque la negociación es un instrumento para la puesta en marcha de un esquema de interacción optimizador (y no meramente maximizador), la maximización restringida entraría en juego en el contexto post-convencional. El contexto de la negociación se mantiene como ámbito por excelencia de la maximización directa, pese a que la razón de ser de la negociación es la posibilidad de optimización que surgiría en un contexto regido por el principio alternativo.

[29] Ofrecemos alguna información sobre el Teorema de Arrow más abajo, p. 77, nota número 60 del cap. II.

[30] J.C. Harsanyi, "Approaches to the Bargaining Problem Before and After the Theory of Games" (*Econometrica*, 24, 1956, pp. 144-157), R.D. Luce y H. Raiffa, *Games and Decisions* (Nueva York, Willey, 1957), J.F. Nash, "The Bargaining Problem" (*Econometrica*, 18, 1950, pp. 155-162), F. Zeuthen, *Problems of Monopoly and Economic Warfare* (Londres, Routledge, 1930).

ción racional. Se trata de un principio que rigurosamente fiel a la racionalidad individual directamente maximizadora que justifica la cooperación: sólo un esquema cooperativo óptimo e igualmente beneficioso para todos sería aceptado por negociadores racionales plenamente informados. El principio dice que "*la cooperación es racional si y sólo si el resultado de la acción cooperativa proporciona [a todos los agentes] un beneficio relativo igual máximo*"[31]. Este principio es un antecedente remoto de lo que en *MA* será el principio del *beneficio relativo maximin*, por lo que "Rational Cooperation" puede considerarse la primera formulación de una teoría de la negociación racional original de Gauthier[32], que complementa la interpretación de la racionalidad como maximización restringida.

El segundo de los artículos que completa la visión de la racionalidad bosquejada en "Reason and Maximization" es "Economic Rationality and Moral Constraints"[33] ("Racionalidad económica y restricciones morales"). Se trata de un texto que refleja una profundización tanto en el análisis matemático de la racionalidad económica como en la comprensión de 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

---

[31] "Rational Cooperation", cit., p. 57. El sentido del principio es evidente: la negociación entre individuos directamente maximizadores conducirá a una distribución igual del beneficio de la cooperación, con la salvedad de que esa igualdad no se puede medir en términos absolutos (pues el beneficio del que hablamos es una utilidad) sino —como ya veíamos al discutir la crítica de Gauthier al principio rawlsiano de la diferencia— únicamente en términos relativos (la relación entre la expectativa de cada negociador, según su contribución al excedente cooperativo, y la utilidad que ya posee en la situación original). La condición subsiguiente de que este beneficio relativo igual sea el máximo posible viene impuesta por el imperativo racional de la optimización.

[32] La originalidad del modelo de Gauthier no queda en entredicho por el hecho de que, en la práctica, los resultados del mismo coinciden con los del modelo de negociación desarrollado independientemente por Kalai-Smorodinsky. Por otro lado, el principio de la cooperación racional aquí ofrecido será modificado ligeramente en *MA* y, sobre todo, será complementado con el principio de la *concesión relativa minimax*, que supone un refinamiento mayor de la teoría de la negociación racional y un distanciamiento de todos los demás modelos defendidos por los principales teóricos de la negociación. No obstante, la deducción a partir de la racionalidad individual maximizadora y la aplicación como principio de justicia, presentes en "Rational Cooperation" sirven de modelo para la discusión sobre la negociación en la obra principal de Gauthier.

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

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 amplia 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 basado en tal racionalidad —la Teoría de la Decisión Racional y Teoría de Juegos— habían quedado ya definitivamente fijados. Quedaba 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 descripción de la racionalidad individual. Además, una teoría moral que se concibe a sí misma como una parte, y quizá la más significativa, de la Teoría de la Elección Racional[34], 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 acordar *un solo* marco para la interacción cooperativa. Un marco o estructura cooperativa que, para ser imparcial y mutuamente beneficioso, ha de ser convenido mediante un procedimiento igualmente imparcial, esto es, respetuoso con cada individuo.

El procedimiento para la elección social es uno de los elementos

[34] Cfr. John Rawls, *Teoría de la Justicia*, México, F.C.E., 1979, p. 34.

nucleares de este tipo de teorías sobre la racionalidad[35], ya que su cara normativa se resuelve en procedimentalismo ético, lo que significa que el procedimiento que sirve como modelo normativo o explicativo para la elección social, sirve también —convertido en procedimiento "ideal" de selección de los principios de la justicia— como expediente de justificación moral. Como hemos visto, Gauthier ya se había aproximado al problema del 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, y al bosquejar —en "Rational Cooperation"— una regla para la negociación (el principio de la cooperación racional) utilizable como principio de la "moral de los hombres económicos"[36]. El siguiente paso consistirá en generalizar el modelo, de modo que se irá perfilando como el núcleo de la moral por acuerdo.

7.- El papel 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?"[37] ("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-

[35] 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.

[36] "Rational Cooperation", cit., p. 62.

[37] 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.

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 entenderse legítimamente como una decisión *unánime*. Pero eso es todo; se trata de una simple decisión, no hay contrato alguno. Gauthier, por el contrario, supone que las condiciones iniciales no serían las descritas por Rawls[38] 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 los que regirán la distribución de los bienes sociales (los beneficios de la cooperación) y, como individuos racionales, cada parte pretenderá obtener tanto como sea posible de dichos beneficios. Se trata, por tanto, de un problema de negociación.

Afortunadamente, la teoría de la negociación, que se desarrolló a lo largo de los años setenta, había logrado ofrecer una posibilidad de escapar al Teorema de imposibilidad de Arrow, proponiendo 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. Nash[39], Ehud Kalai y Meir Smorodinsky[40], así como del propio Gauthier[41]. 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"[42] ("Elección social y justicia distributiva"). Allí discute Gauthier el problema que para la elección social representa el Teorema de Arrow y,

---

[38] Cfr. arriba (pp. 14 y ss.), nuestra breve explicación del artículo "Justice and Natural Endowment: Toward a Critique of the Rawl's Ideological Framework".

[39] John F.Nash, "The Bargaining Problem", cit.

[40] E. Kalai y M. Smorodinsky, "Other Solutions to Nash's Bargaining Problem", *Econometrica*, 43, 1975, pp 513-518.

[41] Nos referimos al comentado "Rational Cooperation" (*Noûs*, 8, 1974, pp. 53-65), que ofrecía un modelo de negociación cuyo resultado, como sabemos, coincidía con el de Kalai-Smorodinsky.

[42] En *Philosophia*, 7, 1978, pp. 239-253.

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"[43]. El artículo defendía y justificaba, por tanto, el papel (político) de la negociación como medio de decisión social[44].

Desde ese momento, Gauthier se esfuerza en mostrar el papel de la negociación racional, no ya en una teoría de la decisión social, sino en una teoría moral. Dos artículos (junto con el capítulo V de *MA*) son fruto inmediato de ese esfuerzo. El primero, "Bargaining our Way into Morality: A Do-It-Yourself Primer"[45] ("Negociar nuestra entrada en la moralidad: manual para principiantes"), es una explicación relativamente sencilla 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

---

[43] Gauthier, "Social Choice and Distributive Justice", cit., p. 252.

[44] 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.

[45] En *Philosophical Exchange*, 2, 1979, pp. 14-27.

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. "Bargaining and Justice"[46] ("Negociación y justicia") debate no sólo con los modelos de negociación de Nash y Kalai-Smorodinsky, sino también con las axiomatizaciones de Roth[47], 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. Pero, como decimos, este artículo representa un pensamiento maduro del autor, y merece una consideración más detenida en otro lugar.

8.- 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 moralidad 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

[46] en *Social Philosophy and Policy*, 2, (1985), pp. 29-47.

[47] Alvin E. Roth, *Axiomatic Models of Bargaining*, Berlin, Springer Verlag, 1979.

concebido por los economistas clásicos. Aparece en el artículo "No Need for Morality: The Case of Competitive Market"[48] ("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 perfecta 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 mejor estado posible.

---

[48] en *Philosophical Exchange*, 3, 1982, pp. 42-55.

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, que analizaremos (junto con los anteriores) al exponer sistemáticamente la teoría moral por acuerdo, no se aclararon hasta los últimos estadios de elaboración de la teoría, como el propio Gauthier confiesa en el prefacio de *MA*. Además, son aspectos tan esenciales a la teoría que no admiten una explicitación fuera del marco de la misma.

9.- El hallazgo de la moral por acuerdo

Con todo, 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. El hecho de que se trate de una teoría contractualista no es más que la consecuencia de una profundización extraordinaria en la compresión 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 Hume[49]. No nos detendremos aquí en estos aspectos (que podemos llamar colaterales) de su producción bibliográfica, pues dedicaremos el capítulo III a los antecedentes históricos del contractualismo moral, y ese lugar será más apropiado para considerar las opiniones de Gauthier sobre autores clásicos. De momento 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"[50] ("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 o 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

[49] Destaca, por lo polémico que resulta, su artículo "David Hume, Contractarian", *Philosophical Review*, 88 (1979), pp. 3-38.

[50] En *Philospphy and Public Affaires*, 6, otoño 1977, pp. 130-164.

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.

10.- 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 misma[51].

En España la obra de Gauthier fue recibida principalmente en el ámbito ius-filosófico, familiarizado con sus antecedentes neo-contractualistas y también con ciertos aspectos de la teoría de la decisión racional[52]. Predominó, entre quienes reflexionaron sobre 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 aciertan a incidir en los aspectos más problemáticos de la teoría de Gauthier, pero tal vez desechan demasiado pronto su proyecto, sin advertir la potencia filosófica del mismo. Esta carencia puede imputarse al hecho de que no ha habido en España una lectura verdadera-

---

[51] 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, 19988); 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 from Morals by Agreement* (Ann Arbor, University of Michigan Press, 1993).

[52] Los comentarios más relevantes a *Morals by Agreement* en español se encuentran en *Doxa*, nº 6 (1989), que dedica a la obra de Gauthier la parte monográfica del número, recogiendo artículos de R. Zimmerling, M. D. Farrell y A. Calsamiglia, y en la *Revista de Filosofía*, 3ª época, vol. IV (1991), nº 5, que añade una interesante recensión de J. Montoya; así mismo, 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).

mente filosófica de la obra de Gauthier. 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, justamente, lo que debemos demostrar en lo que sigue.

Capítulo II

Presupuestos del contractualismo moral

El contractualismo es uno de los caminos posibles para justificar racionalmente las obligaciones morales. El sólo hecho de intentar tal justificación presupone ciertas convicciones sobre el alcance de la razón en materia práctica. Estas convicciones no son, desde luego, incontrovertibles; ni compartidas por todos los miembros del gremio de los filósofos morales. Ya entre quienes participan de la esperanza de hallar un fundamento racional para la moral, no todos consideran que el método contractualista sea correcto (ni, mucho menos, el único correcto). Por esto, es legítimo esperar que el contractualista moral reafirme y defienda filosóficamente sus convicciones básicas (aquellos conceptos nucleares, o postulados, sobre los que se asienta su argumento) antes de presentar su teoría.

La explicitación y defensa de los presupuestos del contractualismo moral adquiere mayor relevancia si tenemos en cuenta que, en esencia, el *modus operandi* del argumento contractualista es reproducido (con mayor o menor sofisticación) desde Hobbes hasta Gauthier, por lo que gran parte de las diferencias en los resultados provienen únicamente de la distinta comprensión de las premisas iniciales. Si consideramos las dos teorías contractualistas de la moral más pregnantes y debatidas en la actualidad, las de Rawls y Gauthier,

hallamos que el origen de sus divergencias se retrotrae al distinto enfoque de los conceptos primordiales. Así, la aclaración de estas concepciones cumplirá, en este trabajo, un doble papel: por un lado mostrará, a grandes rasgos, cuáles son los presupuestos del contractualismo en general; por otro, precisará los caracteres distintivos que toman los mismos en el contractualismo de Gauthier, de modo que resalten su singularidad.

Existe una razón adicional para atender a los supuestos y postulados filosóficos del contractualismo moral. Se trata de que, como el argumento contractualista ha presentado desde sus primeras formulaciones una apariencia de solidez lógica, las críticas más vehementes contra el mismo se han dirigido a socavar sus fundamentos conceptuales. De hecho, quizá la característica más sobresaliente del neocontractualismo, frente al contractualismo clásico, sea el elevado grado de refinamiento que ha alcanzado la descripción y defensa de conceptos como la racionalidad individual, el auto-interés (*self-interest*) o el desinterés mutuo (*mutual unconcern*). Este esfuerzo teórico tiende, sin duda, a minimizar los efectos de las críticas más previsibles; pero posee, además, la virtualidad de ofrecer a quienes comparten la ideología básica del contractualismo un marco apropiado para el debate, pues establece con toda precisión la naturaleza y función de las premisas conceptuales comunes. Por ello, nosotros confiamos en que el estudio que presentaremos inmediatamente beneficie las discusiones posteriores.

En cuanto al contenido de este capítulo, debemos reconocer que no es fácil determinar hasta dónde conviene que llegue la aclaración conceptual y metodológica preliminar: si es demasiado extensa, se adentra en aspectos cuya especialidad exige considerarlos como parte de la teoría, en vez de como premisas; si, por el contrario, nos limitamos a una escueta enunciación del punto de partida ideológico de las teorías contractualistas, no habríamos aportado gran cosa a la comprensión de las mismas. En relación al número de conceptos que deben considerarse fundamentales, nos encontramos ante un dilema semejante: el exceso nos llevaría a analizar innumerables conceptos que el contractualismo asume sin discutir explícitamente —como los de persona,

capacidad de obrar, capacidad de obligarse, sociedad, actuación conforme a reglas, racionalidad, utilidad, maximización, etc.— lo cual nos alejaría, no sólo del objetivo de nuestro trabajo, sino del de las propias teorías contractualistas en general —el que estén basadas en (o sean consecuencia de) una determinada ontología y antropología, no implica que su estudio deba incluir necesariamente el de estas últimas en toda su extensión. Pero si tematizamos muy pocas ideas básicas corremos el riesgo de que algún aspecto realmente crucial para la comprensión de la teoría quede sin analizar.

El modo que nos ha parecido más apropiado para eludir estos problemas de método consiste en tomar como guía las propias palabras de Gauthier quien, en dos párrafos muy significativos del capítulo I de *MA* escribe lo que sigue:

> "La moral por acuerdo ofrece un fundamento racional para distinguir lo que está o no está permitido. Los principios morales se introducen como el objeto de un acuerdo *ex ante*, completamente voluntario, entre personas racionales. Tal acuerdo es hipotético, pues supone un contexto pre-moral para la adopción de reglas y prácticas morales. Pero las partes de ese acuerdo son individuos reales concretos, distinguibles por sus capacidades, situaciones e intereses. En la medida en que acordarían restringir sus decisiones, limitando la persecución de sus intereses particulares, reconocen una distinción entre lo permitido y lo no permitido. Como personas racionales que comprenden la estructura de su interacción, reconocen un lugar para la restricción mutua, y por tanto para una dimensión moral en sus asuntos.
>
> "Por supuesto, hay que demostrar que exista un fundamento racional contractualista de la moral. Esa es la tarea de nuestra teoría. Y nuestro propósito inmediato es relacionar la idea de tal fundamento racional con la introducción de distinciones morales fundamentales. No se trata de un proceso mágico; la moral no sale de un sombrero vacío, como el conejo del ilusionista. Al contrario, nos proponemos defender que surge bastante simple-

mente de la aplicación de la concepción maximizadora de la racionalidad a ciertas estructuras de interacción."[1]

En estos párrafos —donde se mezclan ideas previas y un avance de las conclusiones de la teoría— merecen ser subrayados los siguientes conceptos: "las partes" (definidas a continuación como "individuos reales concretos"), la "concepción maximizadora de la racionalidad", las "distinciones morales fundamentales" y cierto "contexto pre-moral". Estas son las premisas esenciales del argumento contractualista de Gauthier. Más en concreto, el individualismo metodológico y la racionalidad directamente maximizadora de las partes son los dos supuestos clave en las que centraremos nuestro análisis. Desde luego, tales supuestos no agotan los que un jurista llamaría "elementos del contrato" y, en efecto, aunque no se trate de un contrato jurídico, los "elementos" del contrato social son algunos más. Pero al pretender justificar la sociedad, el Estado o, como en nuestro caso, la moral, apelando a un contrato hipotético, el punto de partida *debe* excluir conceptos derivados de instituciones sociales, sistemas políticos o sentimientos morales. En esa medida, el contractualista debe comprometerse con una radical economía teórica. En realidad, en tanto que empresa de reconstrucción racional, la única premisa del contractualismo debería ser la propia racionalidad[2]. Añadimos el individualismo porque es un

[1] *MA*, p. 9.

[2] La idea de que la teoría moral consiste en una "reconstrucción racional" —idea recurrente en los textos de Gauthier— proviene de una de las más preclaras mentes de este siglo en lo que concierne a la metodología de las ciencias sociales; se trata de K. Popper quien, en *La miseria del historicismo* (Madrid, Taurus, 1961, p. 171-172) escribe: "Me refiero a la posibilidad de adoptar en las ciencias sociales lo que se puede llamar el método de la *construcción racional* o lógica, o quizá el 'método cero'. Con esto quiero significar el método de construir un modelo en base a una *suposición de completa racionalidad* (y quizá también sobre la suposición de que poseen información completa) *por parte de todos los individuos* implicados, y luego estimar la desviación de la conducta real de la gente con respecto a la conducta modelo..." (subrayado mío). Aunque el transplante de este método a una teoría moral tal vez no fuera previsto por Popper, esto es precisamente lo que ha intentado el neo-contractualismo. Nuestro objetivo en el presente capítulo, tal como venimos explicando, consiste en aclarar las dos suposiciones (completa racionalidad e individualismo) que están en la base de dicho método.

supuesto intrínseco al modelo contractualista[3]: el contrato como expediente justificador de la sociedad o de la obligación política no habría sido posible sin una concepción individualista previa (más o menos explícita) de las relaciones sociales[4].

Hemos reducido, por tanto, deliberadamente el número de los conceptos nucleares que vamos a tematizar. Evitaremos con ello explicaciones sobre conceptos que, o bien admiten una definición sencilla sobre la que podemos suponer acuerdo (por ejemplo, el concepto de decisión), o bien están contenidos en otros conceptos mayores, y se explicarán al dar cuenta de éstos (por ejemplo, los conceptos de independencia y autonomía, que están incluidos en la idea de individualidad). También queremos enfatizar, con esta reducción, el hecho de que el contractualismo toma muy en serio la demanda filosófica de razonar partiendo del menor número posible de premisas, la cual habría quedado oscurecida si hubiésemos desglosado en exceso cada una de ellas. Sin embargo, la razón decisiva en favor de nuestro planteamiento es que estamos convencidos de que, en efecto, el contractualismo *logra* argumentar desde premisas sumamente sencillas y plausibles y, si en ocasiones parece lo contrario, ello se debe a la confusión entre los puntos de partida necesarios para un argumento, y sus conclusiones.

Como ejemplo de esta confusión tomemos la idea de individuo: es posible desarrollar un concepto muy completo de sujeto moral desde un punto de vista contractualista. Tal concepto formará parte de las conclusiones de la teoría moral que expondremos y, como tal, será discutible —bien en el marco de la teoría, bien mediante la negación de las premisas en que se basa—, pero no debe ser confundido con la premisa argumental misma. Ésta —cifrada en el carácter individual y libre de las partes del contrato— no tiene nada que ver

[3] Más abajo nos referimos al carácter exclusivamente metodológico de este supuesto. Entendemos que este carácter justifica nuestra afirmación anterior sobre la racionalidad.

[4] Cfr. Laurent, A., *Histoire de l'individualisme*, París, Presses Universitaires de France, 1993, cap. II, 3., en relación a la importancia del movimiento político de los "levellers" (niveladores), que profesaron, allá por 1640, un individualismo social absoluto e influyeron en las ideas políticas de Hobbes.

con un concepto de agente moral (sería ilegítimo que así fuera, pues es el inicio de un argumento que debe hacernos ver cómo surge la moralidad a partir de premisas no morales), sino que consiste más bien en una *hipótesis metodológica* que podría incluso ser negada por las conclusiones de la teoría. Por nuestra parte, nos esforzaremos en evitar la aparente complejidad que se deriva de esta confusión frecuente, mediante la restricción al mínimo de los conceptos básicos que someteremos a estudio y la explicación detallada de la interpretación que nos parece correcta.

Ahora bien, incluso reducido el número de las que consideraremos premisas esenciales del argumento contractualista, aún se podría cuestionar el hecho mismo de que les dediquemos un capítulo. La necesidad de explicitar y debatir los supuestos de la teoría parece no ser tan evidente para quienes, como el propio Gauthier, consideran que lo verdaderamente original del contractualismo moral es que "empieza con nuestras concepciones intuitivas de la racionalidad y la moralidad"[5]. Según esta opinión, la explicación del punto de partida sería prácticamente superflua. Sin embargo, si bien es cierto que conceptos como el de "racionalidad simplemente maximizadora" responden sin demasiada dificultad a la idea intuitiva más sencilla sobre la razón instrumental, no es menos cierto que, para hacer de ellos un uso técnico, exigen una compleja elaboración teórica, que recomienda prestarles alguna atención pormenorizada (que es lo que, dicho sea de paso, hacen también Rawls, Buchanan y Tullock e incluso Gauthier). Por otro lado, a nadie escapa que la aceptación de premisas supuestamente neutras depende de pre-concepciones ideológicas o culturales concretas, hecho en que se apoya uno de los argumentos más empleados por quienes critican las éticas liberales inscritas en el proyecto ilustrado, intentándo invalidar la pretendida universalidad de sus conclusiones[6]. Desde luego, el

[5] Gauthier, D., "Justice and Natural Endowment..." (cit.), p. 150.

[6] Sobre este tipo de críticas al proyecto ilustrado, que podemos denominar en general "contextualistas" o "comunitaristas", véase Bhargava, R., *Individualism in Social Science*, Oxford, Clarendon, 1992, pp. 223 y ss., y Thiebaut, C., *Los límites de la comunidad*, Madrid, Centro de Estudios Constitucionales, 1992, cap. primero ("Contra el liberalismo: Neoaristotelismos y Comunitarismo"), pp. 19-64.

profundizar en las premisas no elimina su contextualismo, pero sí puede mitigar o prevenir la crítica, bien mostrando la generalidad de los conceptos (ya que no su universalidad), bien asumiendo su origen contextual, pero sin derivar de él consecuencias fatídicas para la teoría.

De todas formas, este capítulo no pretende realizar una defensa ideológica del contractualismo como método de la filosofía práctica —tal defensa resulta, de momento, innecesaria, en la medida en que vendrá de suyo al paso mismo de la exposición subsiguiente. Nos limitaremos a explicitar lo más claramente posible el sentido y origen de las concepciones nucleares que deben ser tenidas especialmente en cuenta antes de iniciar la discusión de la teoría. De estos conceptos, unos son las verdaderas premisas del argumento: el postulado del individualismo y el axioma de la racionalidad; otros son hipótesis más o menos directamente extraídas de esas premisas, como la idea de una zona exenta de moralidad; por último, la noción previa de moralidad tiene solamente un papel orientador del despliegue argumental, ya que la misma naturaleza de la teoría prohibiría considerarla como premisa[7].

[7] Respecto a esta última noción, debe quedar claro que el hecho de proponer cierta pre-concepción de la moralidad no invalida el objetivo de la teoría (que es fundar las obligaciones morales a partir de premisas no-morales). Se trata de una idea previa meramente tentativa, cuya aceptación condicional es la razón que permite el intento teórico de ofrecer una fundamentación racional de la moral. Pero el resultado de la teoría no queda condicionado por esta idea previa: la moral que el contractualismo justifique, si es que justifica alguna, puede no tener nada que ver con esta moral vislumbrada, que orienta su investigación, pero sin prejuzgar su resultado.

1.- El postulado del individualismo

Ya hemos apuntado que existe una relación necesaria entre individualismo y contractualismo. La idea de contrato social se desarrolló, como modelo filosófico-político, en respuesta a las demandas modernas de legitimación racional de la soberanía, la obligación política o la normatividad social en general. El contractualismo ofrece un modelo de legitimación racional del Estado y la Sociedad basado en el consentimiento (histórico o hipotético) libremente otorgado por todos los individuos que forman parte de la comunidad. Obviamente, una empresa teórica semejante presupone la posibilidad de pensar la sociedad o la comunidad política en términos de relaciones (contractuales) entre individuos. Es decir, que el individualismo, entendido como esa posibilidad de reducir lo social a relaciones entre sujetos independientes (y la sociedad a un agregado de átomos individuales), ha de ser previo, al menos lógicamente, a cualquier formulación del contrato social. Sin embargo, como destaca Alan Laurent, "el individualismo no constituye un dato originario de la humanidad"[8], por lo que una correcta comprensión del contractualismo exige tener presente el despliegue histórico del individualismo. En buena medida, el individualismo es una circunstancia socialmente vivida antes que una categoría social o un concepto científico[9]. No obstante, el aspecto del individualismo que nos interesa en relación al neocontractualismo es el individualismo metodológi-

[8] Laurent, A., *Histoire de l'individualisme*, Cit., p. 22.

[9] Helena Béjar ha escrito las palabras más claras que conozco sobre este proceso: "Lo que fuera en el origen un universal sociológico que acompaña la propia condición humana, se transforma en una categoría que muda su contenido y significación con el correr de los tiempos. Así, el espacio detrás de la puerta va adquiriendo una concreción teórica hasta formar parte intrínseca del mapa cognoscitivo de las sociedades humanas", en *El ámbito íntimo. Privacidad, individualismo y modernidad*, Madrid, Alianza, 1992, p. 15.

co, que sí remite a una conceptualización concreta. Por otro lado, una de las características más destacadas del individualismo es su multiplicidad de caras[10], lo que nos obliga a precisar los rasgos propios del individualismo metodológico contractualista (haciendo hincapié, obviamente, en la versión de Gauthier.

Nuestra aproximación al individualismo se basará en los recientes trabajos de Alan Laurent[11] (por lo que se refiere a la historia del individualismo), Rajeev Bhargava[12] (en relación al individualismo metodológico), y Angeles Jiménez Perona[13] (quien realiza un limitado pero agudo análisis del individualismo metodológico en su versión liberal). También tendremos en cuenta las formulaciones metodológicas clásicas desde Comte y Mill hasta Schumpeter y Popper, así como las concepciones concretas de los contractualistas (especialmente Hobbes y Locke) y neocontractualistas (Nozick, Rawls, Buchanan- Tullock y Gauthier).

En relación a esta última referencia, debemos insistir en que, aunque Gauthier titula el último capítulo de *MA* "El individuo liberal", recogiendo, a modo de conclusión, una descripción de las características y capacidades del individuo libre y autónomo, tal como puede entenderse desde una perspectiva liberal y contractualista de la sociedad, juzgamos que esa caracterización presupone toda una teoría moral por acuerdo, por lo que no es legítimo emplear

---

[10] Laurent (*op. cit.*, p. 19 y ss.) recuerda que existe un individualismo "alemán", que destaca la idea de autoconstrucción individual; otro anglosajón, basado en las nociones de "privacidad" y "propiedad"; otro que destaca la igualdad universal (en Francia); existen versiones del individualismo anarquistas (Stirner), democráticas (Durkheim), liberales (Locke), aristocráticas (Nietzsche) o conservadoras (Hayek).

[11] *Op. cit.*

[12] *Individualism in Social Science. Forms and Limits of a Methodology*, Oxford, Clarendon, 1992.

[13] *Entre el liberalismo y la socialdemocracia. Popper y la "sociedad abierta"*, Barcelona, Anthropos, 1993.

su contenido como explicitación de las premisas de la misma teoría[14]. Es cierto que muchos de los rasgos del individuo liberal tal como es concebido en la conclusión del libro de Gauthier pertenecen también al concepto de individuo que se postula como premisa argumental, pero lo que al inicio es hipótesis metodológica —y así debe ser tratado— es propuesto al término como una tesis plausible. El tratamiento que daremos en este capítulo al individualismo, aunque sea inevitablemente deudor de la concepción definitiva que ofrece Gauthier[15], se ciñe exclusivamente al aspecto metodológico.

Comenzaremos con un breve acercamiento al origen histórico del individualismo y a su papel en las teorías contractualistas clásicas, refiriéndonos básicamente al "individualismo" hobbesiano (epígrafes *a* y *b*). Después nos centraremos en las definiciones del individualismo metodológico, procedentes de las Ciencias Sociales y la Economía (epígrafe *c*), para explicar y analizar finalmente el contenido que el mismo adquiere en la teoría moral de Gauthier (epígrafes *d*, *e* y *f*). Este análisis se complementará, a modo de conclusión, con una comparación entre el individualismo gautheriano y el rawlsiano (epígrafe *g*).

a) El origen del individualismo.-

Schumpeter ha gozado el honor de pasar a la historia como acuñador del afortunado término "individualismo metodológico"[16], otorgando así una clara

---

[14] Ya nos referíamos a este problema arriba, cfr. p. 52.

[15] Esta concepción no está contenida exclusivamente en el último capítulo de *MA*, sino que ha sido desarrollada con gran profundidad en algunos ensayos posteriores, especialmente en "Morality, Rational Choice and Semantic Representation", en Frankel Paul, E. *et al.* (eds.) *The New Social Contract*, Oxford, Blackwell, 1988, pp. 173-221; "Value, reasons and the sense of justice", en Frey, R.G. (ed.) *Value, Welfare and Morality*, Nueva York, Cambridge U.P., 1993, pp. 180-208; "Assure and Threaten", *Ethics*, 104 (julio 1994), pp. 690-721.

[16] Cfr. Bhargava, *op. cit*, p. 1, quien, a su vez cita a Machlup F., *Methodology of Economics and Other Social Sciences*, Nueva York, Academic Press, 1978, p. 472.

filiación y fecha de nacimiento al mismo. Nada parecido ocurre con el término "individualismo", ni con el uso científico-social de "individuo". Más abajo, al hablar del individualismo metodológico, habremos de referirnos a las formulaciones modernas originales de estos conceptos, como las de Hobbes o Mill, pero es significativo que ni uno ni otro emplean el término "individuo"; mucho menos "individualismo"[17]. Sin embargo, es evidente que el proceso de individuación, entendido como la progresiva ampliación de las demandas de libertad y de autonomía de los seres humanos individuales, es una constante en la historia de occidente. Especialmente en la historia moderna pues, como escribe Helena Béjar, "el individualismo es un fenómeno que sólo tiene lugar en condiciones de modernidad, es decir, cuando el orden tradicional comienza a disolverse"[18]. En la modernidad, en efecto, el individualismo surge como reclamación política y como forma de vida. Y ello antes de ser conceptualizado como categoría de las ciencias sociales (al igual que aparece como modo de pensar lo social *de hecho* antes de su consideración metodológica o epistemológica).

El inicio del despliegue histórico que denominamos, siguiendo a Laurent, "proceso de individuación", puede rastrearse en Grecia, entre los sofistas y Sócrates[19]. Aunque podría hablarse de "individualismo" como característica del ideal de virtud (excelencia) aristocrático, se trata, en parte, de una pista falsa. El individualismo (liberal y modernamente entendido) es una demanda democrática que incorpora entre sus ideales el universalismo; ello es así incluso aunque en sus momentos emergentes aparezca como un modo de vida exclusivo de cierta clase social (aristocracia capitalista, burguesía). Sin embargo, lo que en la Grecia clásica puede asimilarse al individualismo es un proyecto aristocrático, que se integra extrañamente en el holismo presente en la concepción

[17] Hobbes habla, a lo largo de su obre, de "hombre/s" o "ciudadano/s". Mill se refiere a los "seres humanos" como "elementos separados" dentro de la sociedad (cfr. *On the Logic of the Moral Sciences*, Nueva York, Bobbs-Merrill, 1965, p. 79).

[18] *op. cit.*, p. 16.

[19] Cfr. Laurent *op. cit.*, Cap. 1.1.

de la comunidad política[20]. Por eso, los sofistas y algunas filosofía helenísticas son los únicos ejemplos de cierto "individualismo" *avant la lettre*, en la medida en que pensaron —desde su visión política relativista y democrática (o demagógica)— el individuo como categoría universal.

El verdadero origen remoto del proceso de individuación moderno es la mezcla de ideas de origen religioso judeo-cristiano (la salvación del alma personal, que dota de valor absoluto a cada individuo; la responsabilidad individual, ejemplificada en la idea de culpa, etc.) y estoico (la idea de la razón como guía interior individual; la idea de libertad de pensamiento) que, tras perder su caldo de cultivo helenístico e imperial romano y quedar durante siglos un tanto desarraigadas, fueron lentamente digeridas a lo largo de la Edad Media para fraguar en los siglos XIII y XIV, tal vez como consecuencia de los conflictos internos de la Iglesia. El nominalismo, como substrato filosófico y epistemológico, y la vida de Guillermo de Ockham, como ejemplo empírico, pueden considerarse, entonces, testigos de la primera apertura hacia el

---

[20] Es un lugar común identificar a Platón y Aristóteles como "los filósofos de la *polis*", recuperadores de la concepción tradicional de comunidad que se había corrompido bajo la influencia democrática y sofística. sin embargo, especialmente por lo que respecta a Platón, el modelo de comunidad que establece "olvida" que el ideal clásico de virtud es un ideal "individualista" (si bien, como hemos señalado, es una "pista falsa" del individualismo). El varón virtuoso expresa su virtud en la medida en que se "distingue", se hace único (por sus hazañas, por sus atributos. Su virtud no consiste en respetar reglas, sino es crearlas o en romperlas de modo único, valioso. Frente a él, el individuo democrático, cuyo paradigma es el *hoplita* (soldado de infantería cuya fuerza reside en formar compactas falanges sin desobedecer nunca la dirección de un jefe), se "distingue" por conformarse a reglas públicas y dejar la libertad, la expresión de su individualidad, para el ámbito de sus negocios privados. Insistimos en que el ideal aristocrático es una pista falsa, mientras que el individuo democrático es la único antecedente griego del individualismo, porque el *valor*, y aun el sentido, de las acciones heroicas está indisociablemente unido a la comunidad a la que el héroe pertenece, depende de ella. Por el contrario, el valor, y aun el sentido, de la comunidad política democrática y de sus reglas, depende de su utilidad para el individuo. Esta inversión propiamente moderna se anticipa por un momento en la Atenas democrática; pero fue un momento breve, porque los "maestros recuperadores de la *polis*" se encargaron de conjurar su "amenaza".

Sobre estas ideas (sobre las que, desgraciadamente, no podemos detenernos sin apartarnos demasiado de nuestro argumento) pueden verse, entre otros, John Gray, *Liberalismo*, Madrid, Alianza, 1994 (p. 16 y ss.); E. A. Havelock, *The Liberal Temper in Greek Politics*, New Haven, Yale U.P., 1957; K.R. Popper, *La sociedad abierta y sus enemigos*, Paidos, Barcelona, 1991; E. Lledó, "Aristóteles y la ética de la *polis*", en V. Camps (ed.) *Historia de la ética* (v. I), Barcelona, Crítica, 1988, pp. 136-207; G. Klosko, *The Development of Plato's Political Theory*, Londres, Methuen & Co., 1986; A. Tovar, *Vida de Sócrates*, Madrid, Alianza, 1988, esp. caps. III, VIII y XII.

individuo.

El renacimiento y la reforma completarán esta apertura, realizando, en el norte, el programa nominalista, al convertir a la iglesia en una asociación de creyentes y haciendo pasar al individuo al primer plano[21]. Mientras, en el sur, el individualismo empapa las hazañas artísticas o bélicas en la era de los descubrimientos, pues ya no se trata de logros de un pueblo o una cultura, sino de actos de un artista, un político o un líder militar concreto. Significativamente, A. Laurent fecha en 1492 el nacimiento del individuo. La conquista, sea de nuevos saberes, técnicas, conocimientos, rutas o continentes simboliza la afirmación del individuo, del sujeto responsable de sus actos y de su originalidad.

El triunfo renacentista del individuo produjo, en los siglos XVII y XVIII, consecuencias de, al menos, tres tipos: primero, el esfuerzo filosófico de pensar la individualidad como esencia (representado por Leibniz); segundo, la aparición de movimientos políticos defensores de un individualismo social absoluto (los *Levellers* ingleses) y, por último, la cada vez más extendida comprensión del derecho de soberanía como aquél permiso que el soberano obtiene a partir, bien de un pacto entre *todos* los ciudadanos (*pactum asociationis*), bien de un pacto de los súbditos con él mismo (*pactum subjectionis*). La expresión máxima de esta comprensión fue la teoría del contrato social de Hobbes, y vino acompañada por la aceptación en las ciencias morales del método compositivo-resolutivo de Galileo. La aplicación de este método a las ciencias morales suponía la convicción de que las instituciones sociales podían explicarse por sus elementos simples, así como el hombre en su totalidad puede ser explicado, de modo mecanicista, a partir de sus elementos o partes y las relaciones funcionales entre ellas.

Por tanto, el siglo XVII alumbra, a la vez, una filosofía de lo individual, un conjunto de convicciones políticas que se traducen en una demanda absoluta de independencia y de derechos iguales para cada propietario y un método "científico" para las ciencias morales, que prefigura —y, en algunos casos,

[21] Cfr. sobre este punto Béjar, H., *op. cit.*, pp. 127-129.

como el de Hobbes, ejemplifica con sorprendente precisión— el individualismo metodológico.

A pesar de que podemos considerar que el despliegue del individualismo a lo largo de los siglos XVII y XVIII obedece a una tendencia general en occidente (el acelerado proceso de individuación), Laurent resalta la diferencia fundamental entre el individualismo "continental" y el "británico y norteamericano". Frente al individualismo racionalista abstracto del continente, se desarrolla en las Islas un individualismo concreto, amparado en la filosofía empirista, en el protestantismo anglicano y en el mercantilismo protocapitalista.

En el continente, la filosofía de Leibniz puede interpretarse como la versión metafísica del dogma teológico de la salvación del alma individual. La independencia y la autosuficiencia individual nos convierte en absolutos y durables permanentemente del mismo modo que lo es el universo, también individual. El racionalismo consagra ideales universales (libertad, igualdad, fraternidad) que prestan su contenido a la, también universal, declaración de los revolucionarios franceses. Indudablemente, estos ideales son trasunto de las demandas concretas de una clase emergente, necesitada de libertad y derechos individuales, que había de basar la convivencia política (y su supervivencia) en lazos de solidaridad universal, en vez de seguir confiando en los marchitos lazos feudales. Pero su formalización tiene un tono universal característico, basado en la Razón.

Por el contrario, en Gran Bretaña los *Levellers* profesan, ya en 1640, un individualismo social absoluto[22]. Overton, en su *Arrow Against All Tyrants*, expresa claramente la convicción individualista de que cada ser vivo posee una propiedad que nadie tiene el derecho a usurpar: la propiedad de sí mismo, del "yo"[23]. Para los *Levellers*, la independencia (de la voluntad de otros, princi-

[22] Cfr. Laurent *cit.*, cap. 2.2.; Macpherson, C.B., *The Political Theory of Possesive Individualism*, Oxford, Oxford U.P., 1988 (undécima reimpresión), caps. II y IV, en especial pp. 148 y ss.

[23] Cfr. Laurent, *cit.*, p. 43; Macpherson, *cit.*, p. 153.

palmente), que consiste en ser propietario de la propia persona, es el atributo esencial del individuo. Pero es una independencia construida desde categorías prácticas: la necesidad de emancipación, el hecho de la propiedad, el deseo de libertad política y mercantil, etc. Se diría que el individualismo (con ese nombre o sin él) empapa el esfuerzo emancipador en la política británica sin necesidad de soporte filosófico alguno.

No obstante, como puntualiza Macpherson[24], el individualismo de los *Levellers* carece todavía de la fuerza del individualismo liberal posterior, aquél representado por John Locke, Bernard de Mandeville, Adam Smith y Jeremy Bentham. Estos autores marcan el paso —producido durante la primera mitad del siglo XVIII— desde las demandas individuales de libertad, aún deudoras de cierto compromiso "comunitario" con la sociedad[25], hasta la defensa positiva del individualismo posesivo egoísta como mecanismo que resulta en el beneficio de todos. Esta defensa, cuyo paradigma puede ser *La teoría de los sentimientos morales* (1756), de Adam Smith, habrá de convertirse en el elemento constitutivo fundamental del sujeto moral utilitarista, que hace virtud de su propia libertad e intereses. Recordemos que en los escritos principales de Bentham destaca la confianza en el individuo como juez autónomo de su bienestar, y la concepción del interés común como la suma de los intereses individuales. Bentham representa el triunfo absoluto del individuo —en cierta manera, la ética utilitarista es uno de los puertos de arribada del proceso de individuación— y nos deja, en los albores del siglo XIX, en el umbral de las formulaciones explícitas del individualismo metodológico, que veremos en Comte y Mill.

Sin embargo, el hecho de que el individualismo haya sido, en los países anglosajones, antes una experiencia de vida y una demanda política y sólo posteriormente el fundamento de una teoría ética o metodológica, no significa

---

[24] Macpherson, C.B., *cit.*, pp. 155 y ss.

[25] En este sentido escribe Macpherson (*cit.*, p 156): "Esta visión de la sociedad humana como el bien esencial, y del valor esencial de la vida en comunidad, está diseminada a lo largo de los escritos de los *Levellers*".

que no haya existido un uso metodológico (tal vez inconsciente o, al menos, no explícitamente reconocido) en la filosofía política. El contractualismo es la teoría que mejor representa dicho uso. En efecto, en un contexto político de afirmación de los derechos individuales frente al soberano, un contexto filosófico de afirmación del método lógico-deductivo y un contexto científico en el cual triunfaba el método resolutivo-compositivo y el mecanicismo, es lógico que la ciencia social se viese influida por categorías que hoy denominaríamos individualistas —justamente aquellas que permitieron construir una filosofía política contractualista moderna. Estas categorías estaban ya presentes en Hobbes —cuyos textos ejemplifican aquél nuevo modo de pensamiento filosófico-político (lo que les hace acreedores de nuestra atención); alcanzaron su mayor desarrollo en el contractualismo de Locke y, en general, son comunes a todo pensamiento contractualista posterior. Veamos a continuación el detalle de esa conexión teórica íntima y originaria entre "individuo" y "contrato".

b) Individuo y contrato.-

El nexo necesario entre concepción individualista de los seres humanos y posibilidad de dar una explicación contractual de las relaciones sociales (y, por ende, a otro nivel, de la sociedad misma) fue reconocido por Gauthier en 1977[26], en un ensayo cuyas ideas sobre este punto han sido recogidas por Jean Hampton en su ya clásico estudio *Hobbes and The Social Contract Tradition*.

La tesis de Gauthier puede resumirse, con sus propias palabras, diciendo que "concebir las relaciones sociales como contractuales es suponer que los hombres, con sus peculiares características humanas, son anteriores a la sociedad"[27]. Hay que explicar —como se apresura a hacer el propio Gaut-

[26] "The social Contract as Ideology", *Philosophy and Public Affaires*, 6, otoño 1977, pp 130-164. Posteriormente recogido en Gauthier, D., *Moral Dealing*, Ithaca, Cornell U.P., 1990, pp. 325-354, por donde se citará.

[27] "The Social Contract as Ideology", *cit.*, p. 331.

hier— que la palabra "anteriores" no tiene necesariamente un sentido temporal. Ciertamente, la teoría del contrato social se caracteriza por expresar esta relación de prioridad en términos de tiempo, pero no exige que los hombres reales, tal como los conocemos, existan o puedan existir antes que la sociedad o fuera de ella. Dicho de otra forma, y a salvo de las precisiones que realizaremos a continuación, la teoría del contrato social requiere un individualismo conceptual o explicativo, pero no necesariamente un individualismo ontológico[28].

No obstante, Gauthier y Hampton están de acuerdo en que el individualismo implicado por una teoría contractual plausible va más allá de una concepción simplemente metodológica o explicativa, es decir, va más allá de lo (en principio) estrictamente necesario[29]. Así, centrando el análisis en el modelo contractual hobbesiano, e intentando precisar la tesis que transcribíamos arriba, añade Gauthier:

> "Lo que exige el contractualismo es, en primer lugar, que los seres humanos individuales no sólo puedan, sino que deban ser entendidos aparte de la sociedad. Las características fundamentales de los hombres no son producto de su existencia social. Por el contrario, constituyen las condiciones de la existencia social de los hombres, al ser la fuente de las motivaciones que subyacen a la acción humana en el Estado de Naturaleza, y que se expresan en la hostilidad natural. Así, el hombre es social porque es humano, y no humano porque es social. En particular, la autoconciencia y el lenguaje deben ser tomados como condiciones, no

[28] Empleo la terminología de Rajeev Bhargava, Cfr. su *Individualism in Social Science*, Oxford, Clarendon, 1992; en especial p. 33, donde ofrece su clasificación de las formas de individualismo.

[29] Hay dos razones relacionadas para este "ir más allá"; aludiremos a ambas más abajo: la primera es que la aceptación del individualismo explicativo supone, en quien lo acepta, la asunción (consciente o inconsciente) de ciertas tesis ontológicas y políticas, la segunda es que el propio uso metodológico del individualismo *debe contener* ciertas presunciones que van más allá del método.

productos, de la sociedad."[30]

Este párrafo resume lo que Hampton ha llamado el "radical individualismo de Hobbes"[31]: un individualismo conceptual radical que coincide en el tiempo, como veíamos en el epígrafe anterior, con las demandas políticas de los *Levellers* ("niveladores"), basadas en la (difusa y discutible, como también señalábamos) certeza ontológica de que cada individuo posee una dignidad propia *frente a* los privilegios de ciertos grupos e incluso frente a las necesidades de la comunidad.

Se trata de un individualismo que, por otro lado, evidencia lo que Rajeev Bhargava denomina "atomismo" y "psicologismo"[32], rasgos que él considera propios del "individualismo ontológico", pero cuyo rastro siempre puede ser hallado entre quienes pretenden emplearlo exclusivamente en su dimensión metodológica. Así, Bhargaba afirma que "una metodología sólo puede ser individualista a condición de que tenga elementos residuales de atomismo y psicologismo"[33]. Es decir, que aunque el contractualismo exige únicamente adoptar un punto de vista (o una metodología) individualista para explicar o justificar ciertos hechos sociales, la misma posibilidad de adoptarlo está condicionada por la pre-suposición de ciertas convicciones antropológicas que caen bajo lo que Bhargava denomina "individualismo ontológico"[34].

---

[30] *Ibid.*, p. 332.

[31] Cfr. Hampton, J., *Hobbes and the Social Contract Tradition*, Cambridge, Cambridge U.P., 1988, Cap. 1.2.; en el mismo sentido cfr. Béjar, H., *op. cit.*, p. 30.

[32] Cfr. Bhargava, R., *Individualism in Social Science*, Oxford, Clarendon, 1992, pp. 40-43.

[33] Bhargava, R., *op. cit.*, p. 44.

[34] Bhargava define el individualismo ontológico como aquél punto de vista que niega que existan hechos propiamente sociales. Especialmente, los actos y estados intencionales —a los que, en última instancia, pueden reducirse los hechos sociales— sólo pertenecen a los individuos: "Es innegable que un individuo es un organismo biológico con estados mentales. Por tanto, los hechos sobre un individuo incluyen sus estados físicos, su comportamiento y sus estados psíquicos, tanto intencionales como no-intencionales, y dentro de la gran categoría de los estados intencionales, aquellos dirigidos hacia los elementos naturales no-humanos del mundo, y aquellos que implican a otros. También incluyen las relaciones e interacciones causales y psicológicas, pero se toma como algo dado que el contenido mental de todas las interacciones es aprehendido e individuado por

El análisis de Bhargava a que nos referimos, se ve corroborado por las conclusiones del estudio de Hampton. Ella está convencida de que el individualismo de Hobbes va más allá del método. Aunque cabría una visión estrictamente metodológica del individualismo en el *Leviatán* (obra en la que es posible descubrir la influencia del método resolutivo-compositivo introducido por W. Harvey y Galileo), Hampton arguye que la antropología contenida en el *De Cive* elimina cualquier duda[35]. Hobbes creyó, en efecto, que las características que hacen humanos a los hombres nos pertenecen desde el nacimiento. Admite, desde luego, que los hombres tienen propiedades interactivas (como el lenguaje) y funcionales (su papel en la sociedad) necesariamente ligadas a su ser-social, pero no cree que sean fundamentales o constitutivas de nuestra naturaleza como personas. Según Hobbes —siempre de acuerdo con la interpretación de Hampton— y otros contractualistas clásicos[36] la socialidad que exhibimos posee un valor instrumental para los hombres en cuanto individuos interesados: la individualidad funda nuestra socialidad; no al revés[37].

Si este es el resultado del análisis del contractualismo hobbesiano, apenas es necesario mencionar el de Locke, que profundiza en el carácter propietario del individuo y da por supuesta su coraza de derechos naturales.

Debemos concluir, sin embargo, volviendo al auténtico sentido del individualismo en la teoría política contractualista. Esta vuelta nos situará en la correcta perspectiva pues, frente a interpretaciones apresuradas que ven en

---

individuos" (*Individualism in Social Science, cit.*, p. 45).

[35] Cfr. *De Cive, English Works*, II, p. XVI.

[36] Hay que advertir que en este punto el neocontractualismo no sigue, en general, al contractualismo clásico. Cfr. en especial Gauthier, D., *MA*, pp. 333-338.

[37] Sobre este punto, Gauthier precisa que el hecho de que la sociedad tenga un fundamento contractual y sea instrumental desde el punto de vista individual no significa que sea arbitraria (por convencional). Gauthier sostiene que si la convencionalidad de la sociedad se basa en la naturaleza humana, no implicará arbitrariedad alguna. De modo que el contractualista "debe mostrar que la sociedad es una expresión indirecta, racionalizada, de características humanas naturales y esenciales" ("The Social Contract as Ideology", *cit.*, p. 334). El contractualista debe mostrar que aunque la socialidad sea instrumental desde el punto de vista individual, esa instrumentalidad responde a un carácter esencial del ser humano, con lo que la sociedad puede tener una justificación racional que supere el mero convencionalismo.

el individualismo de Hobbes (y especialmente en el de Locke) un antecedente del individualismo liberal, hay que recordar que los fuertemente individualistas presupuestos antropológicos están al servicio (en Hobbes como en Locke[38]) de la justificación de un Estado o una sociedad civil que, precisamente porque es el más valioso instrumento al servicio de los individuos, ha de mantener su supremacía sobre ellos. En definitiva, el individualismo (con sus implicaciones políticas y teóricas sobre la libertad, derechos, intereses y racionalidad individuales, instrumentalidad de lo social, etc.) sustituye al derecho divino como expediente legitimador de la soberanía estatal pero, una vez legitimada, ésta mantiene (incluso acrecienta, gracias a esta nueva base racional) su poder absoluto sobre los intereses particulares. Ello resulta más patente en el caso de Hobbes, que sirve así como ejemplo palpable de cómo el postulado individualista sirve para justificar conclusiones holistas. Por tanto, el sentido filosófico-político del individualismo —independientemente del reconocimiento de los derechos individuales de ciertas clases sociales y del entusiasmo con que fuesen recibidos entre los filósofos los cambios hacia la sociedad de mercado, e independientemente también de la antropología abrazada por cada pensador— queda circunscrito a una metodología. Así, cuando Jean Hampton escribe, "de hecho, concluiré afirmando que el argumento del contrato social presupone, cuando menos, un moderado individualismo. Y la medida en que los modernos filósofos políticos queramos usar este método de argumentar para justificar nuestras conclusiones políticas depende de si podemos abrazar el individualismo a él inherente o no"[39], se refiere, sin duda, a un individualismo explicativo, rodeado sólo de las implicaciones ontológicas más débiles e inevitables (atomismo y psicologismo residuales, como resalta Bhargava). Por todo ello, la más precisa definición del individualismo que abraza el contractualista moral

[38] Respecto a Locke, esta afirmación pudiera parecer menos ajustada. Así, Béjar contrapone el discurso hobbesiano, que es individualista en sus premisas y holista en su conclusión, al lockeano, "individualista tanto en sus premisas como en sus conclusiones" (*op. cit.*, p. 36). Es cierto que Locke desacraliza el Estado y elimina el carácter voluntarista del Leviatán, pero a costa de sacralizar la "sociedad civil", a la que el individuo queda igualmente subordinado. En favor de esta postura (y, por tanto, frente a la afirmación de Béjar), véase Macpherson, *op. cit.*, p. 255.

[39] Hampton, J., *op. cit.*, p. 11.

no la hallaremos en la política práctica ni en la teoría social, sino en la metodología de las ciencias humanas.

c) El individualismo metodológico.-

J.S. Mill expone sistemáticamente (y defiende) por primera vez el método individualista empleado, como hemos visto, desde Hobbes. Fue en 1843, en el libro sexto de su sistema de lógica[40]. Su exposición sobre metodología de las ciencias sociales no carece, sin embargo, de antecedentes. El más influyente de ellos es el *Curso de Filosofía Positiva* de A. Comte, como reconoce el propio Mill[41]. Pero ni siquiera la obra de Comte alcanza la precisión y radicalidad de la lógica de Mill. De hecho, es justamente famosa la circunstancia de que éste eliminó de las últimas ediciones del *Sistema de Lógica* muchas de las referencias a Comte (en su mayoría laudatorias), al reconsiderar la exactitud del pensamiento de quien había ejercido sobre su metodología de las ciencias humanas la influencia más significativa.

Mill procede, en su *Lógica de las ciencias morales*, a comprobar qué método científico se ajusta mejor a la naturaleza de los fenómenos sociales. Su conclusión es que gran parte de los errores en la comprensión del objeto de las ciencias humanas se deben a la aplicación de los métodos de la química (método experimental) o de la geometría (método abstracto). Por contra, encuentra que el método de la física (deductivo concreto) es más adecuado a la naturaleza de dicho objeto. El método consiste en el avance de hipótesis y su posterior verificación empírica; con la certeza de que cada efecto se debe a una causa o a una composición de causas, cuya total discriminación se resolverá en el conocimiento (y capacidad de determinación) de las relaciones exactas entre causas y efectos mediante la enunciación de leyes generales. En el ámbito de las ciencias sociales, el método compositivo es de aplicación, porque la

[40] Mill, J.S., *On the Logic of Moral Sciences*, Nueva York, Bobbs-Merrill, 1965.

[41] Mill, J.S., *op. cit.*, p. 81.

sociedad es concebida, al igual que la naturaleza, como un agregado de elementos poseedores de sus propias leyes necesarias:

> "Las leyes de los fenómenos de la sociedad no son, ni pueden ser, otra cosa que las leyes de las acciones y pasiones de los seres humanos reunidos en el estado social. Los hombres siguen siendo hombres aunque se hallen en un estado social; sus acciones y pasiones obedecen a las leyes de la naturaleza humana individual. Los hombres, cuando se reúnen, no se convierten en otra clase de substancia con propiedades distintas; como el hidrógeno y el oxígeno son distintos del agua, o como el hidrógeno, el oxígeno, el carbono y el ázoe son distintos de los nervios, músculos y tendones. Los seres humanos no tienen en sociedad más propiedades que aquellas que se derivan, y pueden resolverse en, las leyes de la naturaleza del individuo. En los fenómenos sociales, la composición de causas es la ley universal."[42]

Este texto, procedente del capítulo 7 de la *Lógica de las ciencias morales*, marca el momento en que Mill, tras haber estudiado las leyes del carácter y acción humanos en los capítulos previos, reconoce que esas mismas leyes (que rige el comportamiento individual), son el fundamento del método apropiado para el hallazgo de las leyes sociales. Aún más claro es el siguiente texto, que además precisa los límites del método individualista:

> "Por muy complejos que sean los fenómenos, todas sus secuencias y coexistencias resultan de las leyes de los elementos separados. El efecto producido, en los fenómenos sociales, por cualquier conjunto complejo de circunstancias equivale precisamente a la suma de los efectos de las circunstancias tomadas singularmente; y la complejidad no proviene del número de leyes, que no es especialmente grande, sino del extraordinario número

[42] Mill, J.S., *op. cit.*, p. 59.

> y variedad de los datos o elementos —de los agentes que, obedeciendo a ese pequeño número de leyes, cooperan para producir el efecto."[43]

Lógicamente, este individualismo (o atomismo) metodológico presupone la "doctrina de la necesidad filosófica", esto es, la tesis de que las conductas y caracteres individuales están determinadas por leyes. Tal doctrina ha de defenderse frente a quienes proclaman la absoluta libertad o indeterminación de la voluntad humana tanto como frente a los fatalistas. Estos últimos, llevando al extremo el determinismo, sostienen que ni nuestro carácter ni nuestro comportamiento puede ser modificado jamás por nuestra voluntad, por lo que ésta se halla presa de leyes y circunstancias ajenos. Entre ambos peligros, Mill trata de definir su idea de necesidad:

> "Correctamente concebida, la doctrina llamada Necesidad Filosófica es simplemente esto: que, dados los motivos presentes en la mente de un individuo, y dadas también las disposiciones y carácter del individuo, la manera en que actuará puede inferirse sin error; que si conociéramos completamente a la persona y todos los condicionamientos que actúan sobre ella, podríamos predecir su conducta con la misma precisión con la que predecimos un evento físico."[44]

Frente a la interpretación fatalista de esta necesidad, que deduciría de ella la imposibilidad (para cada individuo) de pensar o sentir de modo diferente a como lo hace, Mill recuerda que el hecho de que el carácter individual nos venga dado, no significa que no seamos nosotros mismos uno de los elementos que intervienen en su formación y, en esa medida, tenemos la capacidad de

---

[43] Mill, J.S., *op. cit.*, p. 79.

[44] Mill, J.S., *op. cit.*, p. 10.

transformarlo[45].

Tanto el individualismo que presenta Mill, enmarcado en una epistemología de las ciencias sociales y tematizado metodológicamente, como la reflexión sobre el sentido de la necesidad (o determinismo) respecto a las leyes psicológicas individuales y a la formación del carácter, son el antecedente lejano de los supuestos individualistas del moderno contractualismo. En particular, algunas de las posiciones sobre la correlación entre necesidad y libertad individual son esencialmente idénticas a aquellas que presiden, por ejemplo, la idea de "restringir la maximización", propuesta por Gauthier[46]. De hecho, Mill afirmaría que el componente esencial de la individualidad reside en la existencia de una voluntad libre, elemento imprescindible para las teorías del contrato[47].

Sin embargo, la formulación más precisa del individualismo metodológico (y aun la invención de su nombre) tuvieron lugar no entre filósofos, sino entre economistas. Como mencionábamos arriba, se atribuye a Schumpeter la acuñación de tan afortunado término, cuyo referente sería "exclusivamente una estrategia científica de acuerdo con la cual, en la descripción de ciertos procesos económicos, es preferible empezar por las acciones de los individuos"[48]. Felizmente, contamos con definiciones más precisas que esta de Schumpeter, porque él no fue sino uno de los miembros de la Escuela de Viena, formada por seguidores de K. Menger comprometidos con la renovación de los fundamentos epistemológicos de las ciencias económicas y humanas. A esta escuela debemos un avance formidable en la comprensión del método de las ciencias sociales e incluso de la ética, avance en el que se enmarcan las

---

[45] Cfr. Mill, J.S., *op. cit*, pp. 13-14.

[46] Cfr. especialmente "Reason and Maximization", *cit*, y el cap. VI de *MA*.

[47] Hay que precisar, no obstante, que en Mill la voluntad libre se identifica con la libertad moral: "si lo examinamos de cerca, encontraremos que esta sensación de que somos capaces de modificar nuestro propio carácter si queremos, es el mismo sentimiento de libertad moral del que somos conscientes" (*op. cit.*, p. 15). Por el contrario, el contractualista moral tiene suficiente con identificarla con la autonomía y la capacidad de ordenar un conjunto de preferencias individuales.

[48] cit. por Bhargava, R., *cit.*, pp. 1-2.

investigaciones de Von Mises, Popper y Hayek. La Escuela de Viena restauró el individualismo metodológico ya elaborado teóricamente por Comte y Mill. Se trataba de responder a las amenazas del historicismo, el marxismo y otras formas de anti-individualismo cada vez más influyentes en el viejo continente.

De los miembros de la Escuela de Viena, Hayek se caracterizó por su radical defensa de un individualismo (social y político) liberal, inspirado en la tradición anglosajona (Mill, Locke, A. Smith); Von Mises trató de distinguir entre el carácter metodológico del individualismo que defendía la escuela, y lo que llamó "filosofía individualista", asociada con el egoísmo ético. También defendió la neutralidad del método individualista. Pero es sin duda Popper quien de modo más exacto definió, en *La miseria del Historicismo*, el concepto que nos ocupa:

> "La tarea de la ciencia social es la de construir y analizar nuestros modelos sociológicos cuidadosamente en términos descriptivos o nominalistas, es decir, *en términos de individuos*, de sus actitudes, esperanzas, relaciones, etc. —un postulado que se podría llamar 'individualismo metodológico'"[49]

Popper deja sentado el carácter de postulado normativo del individualismo que, como veremos inmediatamente, es precisamente la función que cumple en las teorías contractualistas. El individualismo metodológico, tal como es entendido desde la Escuela de Viena en adelante, se refiere sólo a la posibilidad de una explicación científica de las entidades "sociales", pero se mantiene neutro sobre cuál sea la "realidad" o el "sentido" de esos fenómenos colectivos. Así se entiende la montaraz explicación de John Dunn, quien escribe que el individualismo metodológico "insiste en que ésta es la única manera no supersticiosa en que se pueden comprender esas entidades [sociales], que

[49] Popper, K., *La miseria del historicismo*, Madrid, Taurus, 1961 (trad. Pedro Schwartz), p. 166.

literalmente *ahí* no hay nada más que un condenado individuo tras otro"[50].

Ahora que hemos recogido las formulaciones canónicas del individualismo metodológico, tal vez sea el momento de incidir una vez más en la diferencia entre (empleando palabras de Bhargava) "principios ontológicos *a priori*" y "consecuencias ontológicas de estrategias explicativas afortunadas"[51]. Como hemos visto, el individualismo metodológico se presenta como una alternativa puramente epistemológica: Schumpeter habla de "estrategia científica", Popper de "postulado", etc. Tras el esfuerzo en epistemología realizado por la Escuela de Viena (e independientemente de las interpretaciones más o menos radicales de sus tesis metodológicas entre los "libertarios" o "ultraliberales"), la distinción entre un individualismo político, económico o sociológico y un individualismo propiamente metodológico ha quedado firmemente establecida[52]. Eso no quiere decir que, como ya hubimos de señalar al referirnos a los contractualistas clásicos, tras la adopción del individualismo metodológico no se escondan motivaciones intelectuales que puedan asociarse a una "visión individualista del mundo"[53]. En este sentido, Bhargava parece muy acertado al sostener la debilidad de las dos tesis opuestas sobre la posibilidad de un individualismo estrictamente metodológico, sin supuestos espurios: primera, "es posible distinguir sin ambigüedad los problemas metodológicos de las consideraciones políticas, morales o, más en general, ideológicas y, por lo tanto, el individualismo metodológico puede defenderse sin abrazar el individualismo político y moral"; segunda, "existe una relación interna entre el individualismo metodológico y cierta ideología política,

---

[50] Dunn, J., *La teoría política de occidente ante el futuro*, México, Fondo de Cultura Económica, 1981 (trad. Clementina Zamora), p. 74.

[51] Cfr. Bhargava, R., *cit.*, p. 8.

[52] Sobre esta distinción, ver Jiménez Perona, A., *Entre el liberalismo y la socialdemocracia. Popper y la "sociedad abierta"*, Barcelona, Anthropos, 1993, pp 127-129, y Béjar, H., *op. cit.*, p. 195 y ss.

[53] Cfr. Bhargava, R., *op. cit.*, p. 3.

ética y económica que valora la autonomía, el desarrollo individual, la privacidad y la dignidad inherente a las personas"[54]. En conclusión, por lo tanto, podemos admitir un individualismo puramente metodológico en las ciencias humanas, y en la teoría normativa en particular, pero sin la ingenuidad de creer que su contenido se mantendrá "limpio" de supuestos de otro orden. El compromiso con el individualismo metodológico debe unirse a la cautela para detectar los inevitables supuestos no estrictamente metodológicas en él incorporados.

d) La versión de Gauthier: primeras ideas.-

El contractualismo de Gauthier es, como todo contractualismo, individualista. No se puede negar que Gauthier abraza una concepción individualista del mundo, pues defiende sin rebozo el individualismo liberal, en cuya tradición se inscribe. Sin embargo, tiene buen cuidado de no formular ninguna tesis fuerte sobre antropología filosófica. Ni siquiera en el último capítulo de *MA*, donde tematiza su concepto de un individuo liberal, se atreve a afirmar tajantemente el carácter pre-social o preter-social (o, en otros términos, no-contextual) de las personas concretas. Si esto es así tras la conclusión de su teoría, mucho más lo es al comienzo de la misma. En los primeros momentos Gauthier avanza definiciones que pueden considerarse comprometidas con cierto individualismo político o social, pero muy pronto aparecen las precisiones que aclaran el estátus meramente metodológico que otorga a esas definiciones. Casi a cada paso, Gauthier niega que las características de las personas (de las partes del contrato) que va enumerando representen lo que éstas sean "en realidad", es decir, insertas en la sociedad. El individuo que se toma como parte del contrato es forzosamente un constructo teórico, pues las personas tal como las conocemos poseen características y están en situaciones que harían imposible para ellas adoptar una "posición inicial de

[54] *Op. cit.*, p. 2.

negociación" (Gauthier) o una "situación original" (Rawls). Tal posición —heredera del estado de naturaleza clásico— exige un esfuerzo heurístico que despoja a las personas de sus adjetivos sociales, por lo que la construcción resultante no puede arrogarse el mérito de trasladar una antropología correcta. Se trata simplemente de un postulado metodológico plausible.

Quizá el ejemplo más claro del individualismo contractualista se halle en una obra de Buchanan, en quien Gauthier reconoce que se inspiran algunas de sus ideas[55]. Nos referimos al capítulo segundo ("The Individualistic Postulate") de la obra de J.M. Buchanan y G. Tullock, *The Calculus of Consent*[56]. Este breve capítulo simplemente rechaza cualquier visión organicista de las entidades colectivas (Estado, Sociedad, etc.) y expone las características de los individuos separados que forman las colectividades. Lo más interesante es que deja claro que el individualismo es simplemente un *modo de ver* la acción colectiva como la acción de individuos, cuando deciden realizar sus fines colectivamente, en vez de individualmente[57]. Lo presenta como un postulado normativo, que es exactamente el papel que juega en la teoría moral de Gauthier, e insiste en que

> "el individualismo metodológico no debe confundirse con el 'individualismo' como norma para la organización de la actividad social. El análisis del primer tipo representa un intento de reducir todas las cuestiones de organización política a la confrontación de los individuos con alternativas y su elección entre ellas. La 'lógica de sus decisiones' se convierte en la parte fundamental del análisis, y no es necesario adoptar posición alguna respecto a los fines o criterios últimos que deben dirigir su decisión."[58]

---

[55] Cfr. *MA*, pp. 193 y ss.

[56] Ann Arbor, University of Michigan Press, 1962.

[57] Cfr. Buchanan, J.M. y Tullock, G., *op. cit.*, p. 13.

[58] Buchanan, J.M. y Tullock, G., *op cit*, p. VI.

Buchanan-Tullock y Gauthier comparten asimismo el compromiso de no "introducir ninguna concepción orgánica por la puerta de atrás"[59] como sería el empleo de una "función de bienestar social" que representase "los intereses de la sociedad". El compromiso metodológico exige especial coherencia en este punto: los individuos son las únicas entidades que poseen capacidad de representarse estados futuros que puedan convertirse en fines motivadores de la acción, y los únicos capaces de tomar decisiones basadas en esos fines. Postulado lo cual, la deducción de una función de bienestar social viene prohibida por el Teorema de Arrow[60]. Añadamos que la fidelidad al postulado individualista, unida al Teorema de Arrow, elimina en la práctica la posibilidad de una teoría política o moral utilitarista[61].

Sin embargo, la constante afirmación del carácter metodológico del individualismo contractualista pudiera chocar con el texto de Gauthier que reproducíamos al inicio del capítulo: "...las partes de ese acuerdo son individuos *reales*, concretos, distinguibles por sus capacidades, situaciones e intereses". Un texto posterior se extiende sobre esta misma idea:

> "El individuo real [...] tiene su propias características particulares y definitorias —capacidades, talentos, actitudes, preferencias. Actúa tomándolas como dadas, y su racionalidad se expresa en el esfuerzo por maximizar la satisfacción de sus preferencias dadas sus capacidades y otros rasgos de su carácter en las circunstancias en las que se encuentre, cualesquiera que sean. No hay ningún otro nivel de racionalidad implicado. Ni ninguna otra concepción

---

[59] *Ibid.*

[60] Una definición del Teorema de Arrow puede encontrarse en Resnik, M.D., *Choices*, Minneapolis, University of Minnesota Press, 1993, p. 186. McLean, I. (*Public Choice*, Oxford, Blackwell, 1987, pp. 165 y ss.) ofrece una perspectiva más relacionada con la filosofía política, y añade también la demostración y definición del teorema.

[61] No podemos enumerar aquí las razones de esta conclusión. Digamos simplemente que la toma de conciencia de la misma está en la raíz del resurgir contractualista de este final de siglo, por lo que las razones que aquí omitimos pueden encontrarse en cualquiera de los grandes intentos contractualistas.

> de persona. Concebir la persona de alguna otra forma, por útil que pudiera ser para otros propósitos, sería irrelevante para una teoría de la moralidad basada en la decisión racional."[62]

Estos textos pueden producir cierta perplejidad. Sin embargo, se hacen comprensibles si, saliendo de los límites de *MA*, interrogamos sus puntos de referencia polémicos, la presencia de los cuales explica en buena medida la posición de Gauthier. Hablamos, por este orden, del utilitarismo y de la teoría de la justicia de Rawls[63]. Ambas teorías normativas han caído —junto a una parte de los científicos sociales— en la aceptación de lo que Bhargava denomina un "individuo abstracto":

> "... un individuo teoréticamente posible pero virtualmente no-existente, esencialmente biológico y caracterizado por estados intencionales aislados privados y por un comportamiento a-social."[64]

Este individuo abstracto es tomado como representante de la "individualidad en general", de modo que, subrepticiamente, el postulado individualista se viola, pues en realidad las teorías normativas edificadas sobre tal individualidad abstracta acaban por perder de vista que el elemento primordial de la individualidad consiste en la distinción o independencia entre los individuos, la libertad de sus voluntades, la separación entre ellos, su emancipación de cualquier ligadura que pudiera dar cobijo a una generalización.

Rawls criticó esta violación al escribir que "el utilitarismo no considera

---

[62] *MA*, p. 256.

[63] El individuo liberal, acorazado con sus derechos naturales, tal como lo presenta Nozick en *Anarquía, Estado y utopía* es, por supuesto, otro referente del individuo liberal de Gauthier. Sin embargo, no tiene lugar aquí, pues el individualismo de Nozick es claramente ontológico y moral; viola, por tanto, el límite metodológico que hemos fijado para el análisis (que coincide, obviamente, con el límite del postulado contractualista).

[64] Bhargava, R., *op. cit.*, p. 12.

seriamente la distinción entre personas"[65]. Gauthier va mucho más allá: amplía la crítica al utilitarismo y, además, formula una crítica similar contra Rawls[66]. Respecto al utilitarismo comenta:

> "El utilitarismo no considera seriamente la individualidad de las personas. Nuestra objeción no es tanto que la utilidad de una persona pueda ser sacrificada por la de otras, sino que cada persona es tratada como un medio para la satisfacción de la preferencia total. El utilitarismo viola la integridad del individuo como un ser con sus capacidades y preferencias distintivas."[67]

Ante el peligro de caer en una descripción tan abstracta del individuo que lo despoje de su individualidad, de modo que el acuerdo acabe reducido a un problema de decisión individual (como ocurre en la teoría de Rawls[68]), Gauthier quiere fortalecer la idea de que el postulado del individualismo no significa dejar completamente de lado las condiciones que, de hecho, diferencian a los individuos reales. Las partes del contrato no son individuos socializados, inscritos en un contexto concreto, pero tampoco son unas entidades abstractas que deciden a vista sólo del bien común, u olvidándose por completo de los intereses y capacidades que les definirán como individuos.

Gauthier parece haber asimilado parte de las críticas comunitaristas y habermasianas[69] contra la categorización liberal-ilustrada del individuo-sujeto

---

[65] Rawls, J. *Teoría de la Justicia*, México, Fondo de Cultura Económica, 1979, p. 46.

[66] Cfr. arriba, cap. 1.5.

[67] *MA*, pp 244-245.

[68] Cfr. más abajo sub-epígrafe 'g)', pp. 106 y ss..

[69] Las críticas comunitaristas parecen ser las más radicales, pues "consideran que el individuo se identifica a sí mismo y es identificado por los demás a través de su pertenencia a una multiplicidad de grupos sociales. Soy hermano, primo, nieto, miembro de tal familia, pueblo, tribu. No son características que pertenezcan a los seres humanos accidentalmente, ni de las que deban despojarse para descubrir el 'yo real'. Son parte de mi substancia, definen parcial, y en ocasiones completamente, mis obligaciones y deberes..." (McIntyre, A., *Tras la virtud*, Barcelona, Crítica, 1987, pp. 52-53). Para un resumen de las críticas comunitaristas y neo-aristotélicas al "yo

abstracto, y haberlas empleado como ingrediente de un liberalismo más moderado, pero también anclado en convicciones más seguras y profundas. La reflexión sobre las críticas a un concepto excesivamente abstracto del individuo conduce a Gauthier a esa descripción atrevidamente concreta del individuo-parte del contrato. Con ello Gauthier desafía las cautelas rawlsianas. Sostiene que el velo de ignorancia no es imprescindible, y promete obtener un resultado imparcial sin necesidad de privar a los individuos en el "estado de naturaleza" de la conciencia de que son personas, no sólo con intereses conocidos —recordemos que, tras el velo de ignorancia, los individuos no pueden saber cuáles serán sus intereses— sino también con una identidad concreta y una función de utilidad concreta que representa preferencias privadas sobre la distribución de los bienes sociales. Gauthier piensa que esta profundización en el carácter individual, separado y libre de las personas en la posición inicial no impide la fundamentación de una moral contractual, y sí evita caer presa de las mencionadas críticas al sujeto abstracto, con lo que aumenta la plausibilidad de su teoría[70].

Pero el hecho de que el individuo de Gauthier se aleje del sujeto abstracto del contractualismo rawlsiano no significa que pueda identificarse directamente con las personas tal como las conocemos en la sociedad, porque, paradójicamente, lo que caracteriza a los individuos empíricos es su escasa individualidad: las personas están implicadas en relaciones sociales y afectivas, poseen sentimientos como el altruismo y la solidaridad; y todo ello en medidas

---

liberal", puede verse C. Thiebaut, *Los límites de la comunidad*, Madrid, Centro de Estudios Constitucionales, 1992, p. 47 y ss.; el epígrafe se titula significativamente "No hay un yo sin atributos". En cuanto a la crítica de inspiración habermasiana, se dirige básicamente contra la visión de sujetos iguales, desencarnados y descontextualizados, como la que Rawls presenta al describir las partes en la situación original (pero que no es original de Rawls, sino que reproduce, a grandes rasgos, el individuo utilitarista, el apropiador Lockeano, e incluso el sujeto moral kantiano). La insistencia gautheriana en la "realidad" de las partes en la posición inicial trata precisamente de eludir esta crítica, uno de cuyos frutos más granados es el análisis de Seyla Benhabib en "El otro generalizado y el otro concreto: la controversia Kohlberg-Gilligan y la teoría feminista", en *Teoría feminista y teoría crítica*, Valencia, Edicions Alfons el Magnanim, 1990, p. 119-149.

[70] Sobre esto añadiremos algo en el epígrafe *g)* (p. 53 y ss.) de este mismo punto.

diferentes y relativas, lo que dificulta una aproximación teórica uniforme. Si estos sentimientos y afectos pudieran tomarse como esenciales a los seres humanos, y reducirse a un contenido mínimo auto-evidente, tal vez pudiera fundarse una moralidad sobre ellos. Pero esto no es posible y, además, en la medida en que tales sentimientos se consideran morales, son ellos mismos los que requieren un fundamento racional (si convenimos que la tarea filosófica moderna consiste en ofrecer un fundamento racional de nuestras instituciones morales). Ha de entenderse que con ello no se niega que los sentimientos morales existan ni que sean el motivo inmediato de nuestro comportamiento moral, pero la tarea filosófica de fundamentación exige preguntarse por la posibilidad de su reconstrucción racional, es decir, la posibilidad de quedar justificados ante cada agente racional tomando como único punto de partida la racionalidad meramente prudencial de esos mismos agentes individuales. De este modo, aun tratando de tomar al individuo en toda su complejidad, Gauthier debe despojarlo de los afectos "irracionales"[71] para que funcione como postulado metodológico de su teoría. En los próximos epígrafes vamos a analizar algunas de las características que resaltan en ese individuo despojado de los aditamentos morales inherentes a las personas en la sociedad.

e) Individuos mutuamente desinteresados.-

Debemos a Thomas Hobbes la caracterización más literaria y atractiva de las personas en el estado de naturaleza. De un modo menos elaborado y formal que el usado por los teóricos contemporáneos, Hobbes ya describió al hombre en estado de naturaleza como un individuo, egoísta, interesado sólo en su propia supervivencia, dominado por las pasiones y los deseos e incapaz de mantener la unión y la paz con sus semejantes salvo bajo amenaza. Ya desde

[71] Este adjetivo implica, desde luego, una específica concepción normativa de la racionalidad, que defenderemos, siguiendo a Gauthier, en el punto 2 de este capítulo. Inevitablemente, las concepciones de individuo y racionalidad se solapan y entrecruzan.

entonces, el contractualismo consideró uno de sus méritos (y el primer argumento en su favor) el lograr mostrar cómo surge la sociedad a partir de una situación tan adversa, mediante el correcto uso de la razón.

El contractualista contemporáneo tal vez no exige convicciones antropológicas tan pesimistas, pero sí necesita situar a las partes del contrato en las apropiadas "circunstancias de la justicia", es decir, "aquellas condiciones normales bajo las cuales la cooperación humana es tanto posible como necesaria"[72] por ser beneficiosa para todos. Hume (a quien siguen Rawls y Gauthier en este punto) cifró las circunstancias relevantes de la justicia en a) una moderada escasez y b) conflicto de intereses entre los sujetos. La primera de las circunstancias está en relación con el papel distributivo de la justicia: si un bien es muy abundante, no surgen problemas de distribución, por lo que la aplicación de reglas (justas) para su asignación resulta innecesaria. La segunda está en relación con la caracterización correcta de los individuos en la posición original, y es la que nos interesa ahora. Rawls la explica diciendo: "supondré que las partes en la posición original son mutuamente desinteresadas: no están dispuestas a sacrificar sus intereses en pro de los demás"[73] y, para justificar esta visión restrictiva de la individualidad, añade: "Una concepción de la justicia no debería, por tanto, suponer extensos vínculos de sentimientos naturales. Se trata de hacer la menor cantidad posible de suposiciones en la base de la teoría"[74]. Así pues, el desinterés mutuo o ausencia de "interés en los intereses ajenos" (altruismo) caracteriza al individuo tal como es concebido por la teoría contractualista.

Gauthier ve de modo similar a las partes del acuerdo moral. Los individuos se conciben como *mutuamente desinteresados*. No obstante, Gauthier

---

[72] Rawls, J., *op. cit.*, p. 152. Ver también David Hume, *Treatise*, libro III, part. II, secc. II.

[73] Rawls, J., *op. cit.*, p. 155. Ver también p. 31, donde se puede leer la afortunada fórmula "...se les concibe como seres que no están interesados en los intereses ajenos".

[74] *Ibid.*

considera excesiva la restricción rawlsiana y, siguiendo a P.H. Wicksteed[75], afirma que el desinterés mutuo puede reducirse al "no-tuismo" (*non-tuism*), cuya exigencia es que las preferencias de cada individuo no impliquen al agente con quien interactúa en ese momento, mientras que no importa que incluyan a un tercer sujeto, siempre que no intervenga en la interacción.

Respecto al "desinterés mutuo" —uno de los rasgos principales del postulado individualista del contractualismo— creemos necesario hacer cuatro precisiones (de órdenes diferentes): la primera apunta algunas razones intuitivas en contra de una concepción más "social" del individuo; la segunda distingue el desinterés mutuo del egoísmo; la tercera recuerda las posibilidades del agente así concebido, para evitar que sea identificado con una "máquina de maximizar el propio interés"; la cuarta y última, se refiere al llamado "problema de las coaliciones", un tema conectado con la idea del mutuo desinterés, que nos recuerda la importancia metodológica de concebir individuos independientes.

David Gauthier discute, al inicio de *MA*, un número de razones intuitivas para adoptar el desinterés mutuo como característica de los individuos a partir de los cuales se ha de construir una teoría moral. Desde luego que, en este plano que llamamos "intuitivo" (nos referimos a los fenómenos tal como son experimentados pre-teóricamente por cada uno de nosotros, o tal como nos los explican las ciencias descriptivas, del tipo de la antropología o la sociología), la socialidad es un dato inseparable de la individualidad. Adopta formas conocidas, como la solidaridad con los desfavorecidos, el interés en los asuntos de nuestros familiares cercanos o amigos, etc. En ocasiones, el altruismo y la solidaridad llegan hasta la abnegación, sin merma (y a veces con incremento) de la felicidad del abnegado, pues parten de una decisión voluntaria y libre. Si este tipo de sentimientos o inclinaciones están presentes en nuestra experiencia cotidiana ¿por qué no adoptarlos como elementos característicos de las partes

[75] Cfr. *MA* p. 87. Gauthier sigue, al aceptar el principio de no-Tuismo a Buchanan y Tullock, quienes lo incorporan en el cap. 3 de *The Calculus of Consent* como una de las características del hombre económico. El principio fue formulado originalmente por Philip H. Wicksteed en *The Common Sense of Political Economy*, Londres, Macmillan, 1910, cap. V.

en la posición inicial o estado de naturaleza? En primer lugar, existe una razón de método. Si los sentimientos altruistas fuesen generalizados y muy fuertes, entonces la moral devendría innecesaria, ¿qué necesidad habría de un deber moral si el mayor deseo de cada uno fuese realizar los intereses de los demás, aun a costa de los suyos propios? Pero, al igual que no podemos negar que *hay* sentimientos altruistas, tampoco se puede afirmar que sean generales ni que alcancen tal dimensión que hagan innecesarias las restricciones morales. Más bien lo que ocurre es que los sentimientos solidarios están presentes de modo muy irregular, y tienen normalmente una dirección concreta (se refieren a la propia familia, al propio grupo, etc.). Esto no se niega, pero resulta irrelevante para una teoría que intenta la reconstrucción racional de la moral, y no de una ética concreta de un grupo concreto. Desde este punto de vista, lo adecuado es tomar aquellos caracteres generales que, por mínimos que sean, unifican de alguna forma a las partes en la situación inicial, de modo que permiten proponer un modelo teórico relativamente sencillo[76].

Pero podemos aducir una razón aún más poderosa que estas consideraciones de método. Es una razón deudora de la reciente teoría feminista: se trata de que la sociabilidad y solidaridad características de los seres humanos pueden ser una "fuente de explotación si inducen a las personas a aceptar ciertas instituciones y prácticas que serían costosas para ellas si no fuera por sus sentimientos solidarios. El pensamiento feminista nos ha aclarado esto, que tal vez sea la forma esencial de la explotación humana"[77]. Por eso, la justificación contractualista de la sociedad y la justicia, no puede demandar de los sujetos que acepten instituciones tales que sólo sean beneficiosas para todos *suponiendo* que algunos de ellos sienten simpatía o amor por sus semejantes,

---

[76] En este punto, Gauthier es deudor de Buchanan y Tullock, quienes, tras reconocer la tensión existente entre los sentimientos altruistas y auto-interesados (y la posibilidad de que el predominio de los primeros echase por tierra el postulado individualista), afirman que "en última instancia, la defensa del supuesto del comportamiento económico-individualista debe ser empírica. Si, gracias al uso de este supuesto somos capaces de desarrollar hipótesis sobre la decisión colectiva que ayuden a explicar y comprender las instituciones observables, no es necesario añadir nada más." (Buchanan y tullock, *The Calculus of Consent*, Ann Arbor, University of Michigan Press, 1965, p. 28).

[77] Gauthier, D., *MA*, p. 11.

y están dispuestos a sacrificarse (y a ser felices haciéndolo) por otros. Las instituciones radicalmente justas o imparciales, aquellas que sí pueden reclamar aceptación racional universal han de ser tales que reporten beneficios a todos sus miembros *con independencia* de sus sentimientos hacia otros. Tales sentimientos, cuya existencia no se discute, enriquecen la vida humana una vez establecida sobre instituciones justas e imparciales; pero si (una vez superadas las dificultades de método que apuntábamos antes) se situaran en la base de su reconstrucción racional, podrían legitimar instituciones *esencialmente* injustas[78], como el patriarcado o la esclavitud. Ante este peligro, el contractualista pone entre paréntesis los sentimientos a que nos referimos, desnuda al agente de su "amor al prójimo" y, a efectos metodológicos, lo toma como "mutuamente desinteresado".

Una segunda precisión se dirige contra la fácil identificación entre "desinterés mutuo" y egoísmo. En este sentido, el neocontractualismo debe mostrar muy claramente la distancia que lo separa del contrato hobbesiano. El individuo natural de Hobbes, necesaria y exclusivamente preocupado por su propia conservación, puede aparecer como un egoísta, es decir, como un agente cuyas preferencias se ordenan de tal modo que el estado de cosas más preferido lo es a causa del valor que posee *para* él mismo (se niega que haya valores o utilidades objetivos), y ese valor depende de la medida en que (dicho estado de cosas) representa la satisfacción de deseos o intereses cuyo contenido *se refiere también al* mismo agente[79]. Pues bien, el desinterés mutuo no implica

[78] Empleamos, un tanto libremente, el término introducido por Gauthier en los caps. X y XI de *MA*. Ver, especialmente, *MA* p. 340.

[79] Es importante destacar la diferencia entre "interés *del* agente" e "interés *referido al* agente". sólo en este segundo caso hablamos de egoísmo en sentido estricto. El primer caso incluye a egoístas y no-egoístas (pues las preferencias altruistas también lo son *de* alguien, y los estados de cosas beneficiosos para terceros también satisfacen el interés *del* altruista, sin que ello quiera decir que es egoísta). Con el lenguaje de la economía, esto mismo se puede explicar diciendo que el egoísta sería aquél agente cuya función de utilidad es independiente de las funciones de utilidad de otros *y además* supone un orden de preferencias en el cual los intereses referidos (o dirigidos) a sí mismo ocupan el primer lugar. El altruista "puro" sería, por otro lado aquél agente cuya función de utilidad *reflejase* o *incluyese* la función de utilidad de un tercero. Esta distinción es fundamental, porque entre el altruista (cuya función de utilidad incluye la de otros) y el egoísta (cuya función

necesariamente el egoísmo, tal como queda definido.

Esta es una tesis recurrente entre los teóricos del individualismo metodológico, a causa de las malinterpretaciones de sus críticos. Frecuentemente se confunde el método individualista con la llamada "filosofía individualista", expresión peyorativa que identifica al individualismo con la defensa de cierto egoísmo insolidario. Frente al "individualismo", entendido como egoísmo e insolidaridad, se opone la sociedad colectivista, comunitaria o cerrada. De ahí los esfuerzos liberales para precisar el sentido propio del individualismo[80].

La misma precisión que los teóricos liberales han de hacer frente a quienes malinterpretan el alcance de su individualismo, ha sido reproducida por las teorías normativas contractaulistas de raíz liberal. Así, Rawls escribe:

> "Un rasgo de la justicia como imparcialidad es el pensar que los miembros del grupo en la situación inicial son racionales y mutuamente desinteresados. *Esto no quiere decir que sean egoístas, es decir, que sean individuos que sólo tengan cierto tipo de intereses, tales como riqueza, prestigio y poder.*"[81]

---

de utilidad incluye sólo los intereses referidos a sí mismo) se halla el sujeto mutuamente desinteresado, que puede ser definido, en estos mismo términos, como aquél cuya función de utilidad es independiente de las funciones de utilidad de los demás, pero sin la cualificación que añadíamos en el caso del egoísta (es decir, puede incluir preferenicas *referidas a* otros). Tendríamos, así, el siguiente esquema:

Agente egoísta	intereses del agente (sujeto)	referidos a sí mismo (como objeto)
Agente mutuamente desinteresado	intereses del agente (sujeto)	referidos a cualquier objeto
Agente altruista puro	intereses *de otro/s agente/s* (tomados como sujeto)	referidos a cualquier objeto

[80] En este sentido, podemos recordar las palabras de Hayek, en *Contra la esclavitud*: "el individualismo del que hablamos (...) no está en relación necesariamente con el egoísmo. ¿En qué consiste, entonces, el individualismo? En respetar al individuo por lo que es, reconocer que sus opiniones y sus gustos le pertenecen solamente a él (...), desear que los hombres desarrollen sus capacidades y tendencias individuales" (cit. por Lauren, A., *op. cit.*, p. 102).

[81] Rawls, J. *op. cit.*, p. 31; subrayado mío.

Por su parte, David Gauthier tematizó, en los años setenta y ochenta, el problema del egoísmo[82]. Tras sus análisis, descarta el egoísmo como principio ético y como principio de elección racional. De hecho, la demostración de la incoherencia lógica del egoísmo como principio de la racionalidad práctica en situaciones de interacción —que puede verse en "The Impossibility of Rational Egoism"— conduce a la negación de la posibilidad de un principio ético egoísta, pues el proyecto neocontractualista se basa precisamente en la idea de que las reglas morales son un subconjunto de los principios racionales de elección (en situaciones de decisión estratégica)[83]. No obstante, el punto que conviene resaltar es el siguiente: que aunque un contractualista puede aceptar el egoísmo psicológico —entendido como cierta tesis sobre la motivación humana[84]— e incluso el egoísmo racional —entendido como principio general de decisión— en condiciones de elección paramétrica (y sólo en esas condiciones), reconoce que resulta auto-frustrante (*self-defeating*) cuando se pretende aplicar en condiciones de elección estratégica; y como quiera que el "estado de naturaleza" (especialmente si es entendido como una "situación ideal de negociación") establece condiciones estratégicas, el egoísmo habrá de ser sustituido por algún otro principio de elección que escape a la incoherencia de aquél. Presumimos que, en tanto los hombres en estado de naturaleza son seres

---

[82] Pueden verse "The Impossibility of Rational Egoism", *the Journal of Philosophy*, vol LXXI, nº 14, agosto de 1974, pp 439-456; "The Irrationality of Choosing Egoism: a Replay to Eshelman", *Canadian Journal of Philosophy*, 10 (1980), pp. 179-187; y "The Incompleat Egoist" en McMurrin, S.M. (ed.), *The Tanner Lectures on Human Values*, vol. 5, Salt Lake City, University of Utah Press, 1984, pp. 67-119, reimpreso en *Moral Dealing* (cit.), pp. 234-273.

[83] En un sentido parecido, Rawls niega que el egoísmo pueda siquiera figurar en la lista de principios sobre los cuales los sujetos en la posición original han de elegir. Rawls afirma que un principio egoísta violaría las condiciones formales aceptables desde el punto de vista de la justicia (cfr. Rawls, J., *op. cit.*, p. 162). Gauthier expone su alternativa al egoísmo como principio ético (o de decisión racional en "The Incompleat Egoist" (cit.). No tematizamos aquí esa alternativa, que consiste nada menos que en su propuesta de teoría moral.

[84] Hago mía la definición de egoísmo psicológico de Gregory Kavka (en *Hobbesian Moral and Political Theory*, Princeton, Princeton U.P., 1986, p. 35): "la tesis de que todas las acciones humanas están motivadas por el auto-interés [*self-interest*]". Debemos precisar, no obstante, que, en lo sucesivo emplearemos el tecnicismo auto-interés para referirnos a cualquier preferencia *del* agente (para distinguirlo de unas hipotéticas "preferencias objetivas" o "altruistas". En este sentido, auto-interés no tendrá que ver necesariamente con ningún tipo de egoísmo, sino con el carácter independiente y subjetivo de las preferenicas humanas.

racionales, estarán dispuestos a abandonar el egoísmo e intentar dicha sustitución.

Por tanto, las partes en la situación inicial no pueden ser concebidas como sujetos inevitablemente egoístas, pues ello equivaldría a considerarlos irracionales. No se niega que los sujetos posean motivaciones individuales auto-interesadas; tampoco que algunos de ellos puedan adoptar principios egoístas como guías de su acción, pero tal egoísmo será un rasgo contingente de *algunos* sujetos, pues ni viene lógicamente implicado por el desinterés mutuo, ni se supone necesariamente inscrito en la naturaleza humana (como en el caso de Hobbes). Muy al contrario, es incompatible con la racionalidad individual maximizadora.

Esta reflexión nos conduce a la tercera de las precisiones que anunciábamos sobre el desinterés mutuo. Porque suponer que un agente racional logra superar el determinismo del egoísmo psicológico (en la hipótesis radical de que el egoísmo psicológico refleje la fuente última de la motivación humana) significa dotarlo de una capacidad que el mismo Hobbes no reconoció: la capacidad de dejar de ser una simple máquina de maximizar, para transformarse en un ser apto para suscribir y cumplir los compromisos que inevitablemente se le exigirán si la vida se concibe como un proyecto global —aunque esos compromisos exijan sacrificios ocasionales. En efecto, la necesidad hobbesiana de autoridad y coacción en el estado social tiene su raíz en la idea de que los seres humanos no pueden escapar a sus motivaciones (que se conciben como determinaciones naturales). El contractualismo moral niega que el individuo funcione de un modo tan maquinal. Para los neocontractualistas, el individuo —al margen de sus motivaciones concretas en cada momento y de la motivación global de maximizar su bienestar que se le supone[85]— es capaz de adoptar puntos de vista ajenos a la simple maximización auto-interesada, gracias a la

[85] Volveremos sobre esta afirmación en el próximo epígrafe, pero pueden verse, como ilustraciones paradigmáticas de esta tesis: Gauthier, D., "Rational Choice and Semantic Representation", en Paul, E.F., *et al.* (eds.), *The New Social Contract*, Oxford, Blackwell, 1988, pp. 173-221, en especial p. 219; Gauthier, D. "Assure and Threaten", *Ethics* 104 (Julio 1994) pp. 690-721.

reflexión y a la auto-crítica. En palabras de Gauthier:

> "Podríamos imaginarnos seres tan mecanizados que la maximización directa sería para ellos el único modo de elegir en contextos estratégicos. Tal vez Hobbes pensara que los seres humanos están tan determinados que son simples máquinas de maximizar. Pero si pensó tal cosa, estaba completamente equivocado. La posibilidad de reflexión auto-crítica se halla en el núcleo mismo de nuestra capacidad racional. Un ser plenamente racional es capaz de reflexionar sobre el principio de su deliberación y cambiar ese principio a la luz de la reflexión."[86]

El fundamento de esta perspectiva se halla en que el ser humano no se concibe sólo como un individuo económico (aunque también, como veremos a continuación), sino básicamente como un ser capaz de representarse estados de cosas presentes y futuros y ordenarlos según sus preferencias y según cierto plan coherente de vida, de modo que la motivación no funciona sólo en el nivel de las decisiones concretas, sino también en el nivel de la elección de proyectos de vida globales. Gracias a su capacidad de cuestionar y transformar los principios de su decisión en vista de sus fines globales, el individuo tiene abiertas las puertas de un pacto, negociación o diálogo que pueden fructificar en un acuerdo social y moral para el beneficio mutuo. Que los individuos se conciban como mutuamente desinteresados no excluye que puedan ser egoístas en sus motivaciones (aunque, como veíamos, no lo requiere), pero sí excluye que sean incapaces de escapar a una motivación miope y exclusivamente egoísta. La sujeción exclusiva a tal motivación es irracional.

En cuarto lugar aludiremos brevemente a una cuestión relacionada con un aspecto crucial en la concepción de las partes del contrato, suscitada con posterioridad a *MA*. Se trata del problema de las coaliciones, uno de los tópicos de la teoría de la negociación racional. Gauthier y los neocontractualistas en

[86] *MA*, p. 183.

general (salvo, tal vez, Nozick) heredan la idea de que el "pacto originario" es un acuerdo de cada individuo con todos los demás, o una decisión unánime. Al no discutir esta idea, pasan por alto el hecho de que, definido un estado de naturaleza o posición inicial donde interactúan individuos perfectamente racionales, se producirían de modo espontáneo coaliciones tiránicas; esto es, grupos de individuos que encontrarían más ventajoso pactar entre sí y luego, desde una posición de fuerza, pactar con cada uno de los demás, obteniendo beneficios superiores a los que habrían obtenido mediante un pacto unánime suscrito en una posición de igualdad.

El problema de las coaliciones tiránicas no es en absoluto trivial. Especialmente si tenemos en cuenta que el contrato, tal como lo concibe Gauthier, puede representarse como una negociación racional en la que cada agente negociador actúa persiguiendo maximizar su beneficio. Durante la negociación no hay límite a la maximización de la utilidad individual. Esto significa que cualquier estrategia maximizadora puede ser empleada, y la estrategia de promover y participar en coaliciones es la más eficaz. La consecuencia es que, como afirma Koons[87], "una mayoría cohesionada que actúe en coalición puede imponer términos relativamente desfavorables a los otros agentes racionales de su sociedad". Este resultado es inevitable dada la concepción de las partes y la situación inicial tal como quedan descritas hasta ahora; y es un resultado que echa por tierra el intento de emplear la teoría de la negociación racional como garantía de imparcialidad del contrato moral. Como hemos comentado, Gauthier simplemente pasa por alto este problema en *MA* y, con posterioridad, tan solo se ocupa (brevemente) del mismo en "Moral Artifice", al replicar a Jean Hampton[88].

---

[87] Robert C. Koons, "Gauthier and the Rationality of Justice", *Philosophical Studies*, vol. 76, nº 1 (Octubre 1994), pp. 1-26; p. 22.

[88] Jean Hampton, "Can We Agree on Morals?", *Canadian Journal of Philosophy*, vol. 18, nº 2 (Junio 1988), pp. 331-356. Hampton emplea la noción de "núcleo" (*core*), procedente de la Teoría de Juegos para poner en evidencia el olvido de Gauthier. Según la teoría de juegos, un acuerdo entre jugadores se encuentra en el "núcleo" si el resultado que ofrece es (1) óptimo de Pareto, (2) no puede ser mejorado, desde el punto de vista de cada jugador individual, jugando en solitario y (3) no puede ser mejorado, desde el punto de vista de algún subconjunto de jugadores, dejando el acuerdo y formando una coalición alternativa independiente. La crítica de Hampton consiste en denunciar que Gauthier olvida el requisito tercero: la solución racional de una

Tres posibles soluciones se han presentado al problema de las coaliciones tiránicas, todas ellas muy limitadas y tentativas: la primera fue sugerida por Hampton, y consistía en adoptar un modelo de negociación distinto del de Gauthier. No es este el lugar para analizar la propuesta de Hampton; digamos simplemente que un análisis profundo de la misma mostraría, probablemente, el mismo error descubierto en la de Gauthier[89]. La segunda solución fue apuntada por Gauthier en su replica a Hampton. Gauthier simplemente intenta salvar su teoría mostrando que, en ciertas situaciones-límite, no hay coaliciones tales que mejoren significativamente los resultados de los participantes respecto de los resultados arrojados por la negociación estándar. No obstante, al tratarse de una prueba para casos límite no resuelve el problema general, sino que sólo apunta la dirección en que podría ofrecerse una solución definitiva satisfactoria[90]. La tercera solución, tal vez la más plausible —y la que nos devuelve al problema de la correcta caracterización de las partes del contrato— es la de Robert Koons. Éste parte del hecho de que nada puede garantizar la estabilidad de las coaliciones tiránicas a lo largo del tiempo, luego un agente que se esté beneficiando en el tiempo t_1 de la pertenencia a una coalición dominante, no puede tener la certeza de que en el tiempo t_2 mantendrá dicho beneficio. Ante la posibilidad de estos cambios, los agentes racionales encontrarán preferible erigir estructuras estables e imparciales (basadas en acuerdos unánimes) en vez de permitir la existencia de estructuras beneficiosas para algunos, perjudiciales

---

negociación no sólo depende de lo que podría obtener cada negociador en solitario, sino también de lo que podría obtener adhiriéndose a alguna coalición alternativa. Frente a este segundo punto de referencia, el "contrato social" incumpliría los requisitos de la racionalidad maximizadora.

La replica de Gauthier se basa en la demostración de que, en la mayoría de las situaciones de negociación, el "núcleo" está vacío. Es decir, no existe ningún acuerdo parcial entre jugadores que cumpla las condiciones necesarias para estar en el núcleo; el acuerdo unánime (la solución estándar) es el resultado más próximo al "núcleo". No obstante, en los casos en que el "núcleo" no es vacío, Gauthier reconoce que no es posible alcanzar soluciones definitivas siguiendo su propio modelo de negociación y deja el problema abierto, suponiendo —pero sin poder probarlo— que, tal vez, una aplicación sucesiva del principio de la negociación (el principio de la Concesión Relativa Minimax) arrojaría resultados en el "núcleo" (Cfr. "Moral Artifice", *cit.*, pp. 395-398), con lo que escaparía, en última instancia, a la crítica de Hampton.

[89] La propia Jean Hampton lo admite en las pp. 341-42 del texto citado.

[90] Cfr. Gauthier, D., "Moral Artifice", *cit.*, p. 398.

para otros, y cuyo resultado a medio y largo plazo es imprevisible. Con las palabras de Koons: "si los agentes tienen suficiente aversión al riesgo y no desatienden demasiado los males futuros, puede que encuentren racional dejar pasar las oportunidades presentes de beneficios tiránicos y accedan a una prohibición universal de constituir coaliciones. Esto significaría aceptar estructuras imparciales según el modelo de negociación racional atomístico de Gauthier"[91].

En nuestra opinión, el problema de las coaliciones pone de manifiesto que el individuo (parte del contrato) ha de ser concebido no sólo como un agente desinteresado, sino también como un agente *independiente* que afronta la negociación conducente al acuerdo moral en pie de igualdad con sus semejantes. Igualdad e independencia están garantizadas si se prohíben las coaliciones o, lo que es lo mismo, se supone que los individuos acuden a la negociación de modo separado, defendiendo cada uno de ellos solamente sus intereses particulares. Desde un punto de vista, esta independencia en la negociación puede tomarse como un límite metodológico para garantizar la imparcialidad del resultado. Pero en la medida en que podemos suponer que es racional evitar el riesgo y tener en cuenta los posibles males futuros, la prohibición de las coaliciones tiene una justificación racional no meramente metodológica.

Por último, y a salvo de lo último que hemos dicho sobre la justificación de la prohibición de las coaliciones, queremos insistir en el carácter metodológico del concepto que se va configurando con las explicaciones anteriores —y también las que seguirán—, es decir, el concepto de los individuos en la posición inicial o estado de naturaleza. Tanto Rawls como Gauthier resaltan, especialmente al referirse al desinterés mutuo, que los seres humanos reales poseen evidentes sentimientos sociales, lo cual no es negado por el hecho de adoptar cierto punto de vista metodológico. Insistimos en este punto a fin de contrarrestar la fácil y previsible crítica que Gauthier recoge en la página 100 de *MA*:

[91] Koons, R. C., "Gauthier and the Rationality of Justice", *cit.*, p. 24.

> "El supuesto del desinterés mutuo puede ser criticada por considerarlo generalmente falso o porque, sea verdadero o falso, se piense que refleja una visión excesivamente malévola de la naturaleza humana, destructiva tanto para la moralidad como para los lazos afectivos que mantienen cualquier sociedad humana. Pero tales críticas no comprenden el papel del supuesto."

En efecto, si el papel del supuesto del desinterés mutuo y la independencia en la negociación se entiende correctamente como metodológico, se puede admitir la existencia de lazos de mutuo interés sin por ello tener que introducirlos en la posición inicial. Tales lazos son, por otro lado, muy "particulares y parciales"[92], de modo que —incluso fuera del marco metodológico— tampoco podrían tomarse como característicos del modo en que los seres humanos se orientan hacia los otros. Allí donde terminan las relaciones de sangre y amistad, lo que queda es, generalmente, el desinterés mutuo. El ver esto claramente no supone pesimismo alguno, sino más bien un realismo que hace más plausible el supuesto metodológico contractualista del desinterés mutuo.

Gauthier llega incluso a apoyarse en la historia de las sociedades capitalistas de mercado, como ejemplo fáctico de cómo la socialidad y el beneficio mutuo pueden expresarse en contextos de predominio de las relaciones puramente contractuales y mutuamente desinteresadas (a la vez que particularmente beneficiosas). Sostiene que las sociedades en las que se han desarrollado instituciones capaces de articular los intereses particulares (egoístas) de modo beneficioso para todos, tienen muchas más expectativas de libertad y crecimiento que aquellas otras que siguen estando basadas en instituciones que sólo se mantienen gracias a los sentimientos de solidaridad familiar, tribal, nacional, etc. Aunque el predominio de relaciones contractuales no excluya las relaciones solidarias, la existencia y preponderancia de las sociedades de mercado parece ser una razón en favor de considerar el desinterés mutuo como la característica más general de las relaciones entre los seres humanos, mientras los lazos de sangre o afinidad son una (maravillosa,

[92] *MA*, p. 101.

desde luego) excepción.

Nosotros no creemos, sin embargo, que la apelación a ejemplos fácticos sea necesaria ni conveniente para justificar el supuesto del desinterés mutuo. Estos ejemplos tratan de contrarrestar una crítica que confunde el papel del postulado individualista pasándose al mismo terreno (equivocado) de los críticos. La justificación del postulado no debe conceder esta ventaja a la crítica. Existen razones de método para elegir individuos mutuamente desinteresados como partes del contrato, y estas razones no tienen nada que ver con el hecho de que las personas en las sociedades reales sean más o menos "individualistas", egoístas o narcisistas. Incluso si concediéramos que la vida social sólo puede desarrollarse a partir de ciertos lazos afectivos entre los hombres y mujeres, aún tendríamos que suponer el desinterés mutuo si lo que queremos es fundar una moralidad en los límites (y como parte) de la Teoría de la Decisión Racional. Esta suposición forma parte de un postulado y, aunque está relacionada con caracteres que exhiben los sujetos reales, hacer hincapié en ellos desvirtúa su papel en la teoría. Como tal postulado no requiere confirmación empírica, sino sólo un asentimiento condicional, hasta alcanzar la conclusión del argumento, esto es, la justificación racional de las restricciones morales.

f) Individuo económico y "yo de mercado".-

Para una teoría moral contractualista es fundamental definir un contexto de interacción pre-moral. Tal contexto no se ha dado nunca en la práctica y seguramente no puede darse. Se trata, por tanto, de una ficción, una construcción heurística que sirve para otorgar plausibilidad al método constructivista en la ética: si es posible pensar un contexto racional no-moral, entonces la moralidad sólo estará justificada si puede construirse racionalmente a partir de los elementos pre-existentes es ese estado previo.

Mas el hecho de que el contexto pre-moral del que hablamos sea una ficción teórica no significa que los individuos que toman parte en el mismo puedan ser arbitrariamente caracterizados. Para que el modelo funcione, es

necesario suponer que las partes exhiben cierta regularidad en sus acciones, es decir, que sus actos responden a algún patrón determinado relativamente uniforme[93]. Para que el modelo posea eficacia normativa es necesario, además, que dicha regularidad refleje lo más fielmente posible aquellas características comunes irreductibles que comparten los agentes racionales individuales.

Sería fácil, por ejemplo, llevar a cabo la empresa de fundamentación racional de la moralidad si se postula una "racionalidad" tal que incluyera a la vez, como partes suyas, (1) la presencia evidente de un fin común (digamos, la felicidad de la humanidad), (2) un arsenal de conocimientos sobre los medios adecuados para lograr ese fin y, (3) un sistema infalible de conexión entre la voluntad y la acción[94]. Tampoco sería mal ejercicio escolástico el tratar de hallar principios morales suponiendo que la racionalidad de los agentes fuese completamente dispar y que, por ejemplo, la coherencia entre decisiones y acciones dependiese de una ruleta que gira al azar; o suponiendo que las preferencias no tuviesen nada que ver con las decisiones, de modo que el hecho de estimar mejor para uno mismo cierto estado de cosas no tuviese consecuencias prácticas; o incluso suponiendo que no hay tal cosa como preferencias, o que éstas son completamente aleatorias. Estos ejemplos muestran supuestos implausibles, de los que no se derivarían conclusiones normativas. En el primer ejemplo, por exceso (se presume una racionalidad inflada con contenidos morales); en los siguientes, por defecto, porque no se ajustan a las experiencias

---

[93] Se ha criticado al contractualismo por suponer ilegítimamente esa uniformidad o "igual racionalidad de las partes". Se le imputa que dicha igualdad es un pre-supuesto moral injustificado. Discutiremos este problema más abajo (Cfr. punto 4 de este capítulo, y cap. IV, puntos 3.f y 5.d). De momento, recordaremos solamente las palabras de Hobbes en las que se inspira esta tradición: "La Naturaleza ha hecho a los hombres *tan iguales* en las facultades del cuerpo y la mente, que aunque se halle a veces un hombre de cuerpo manifiestamente más fuerte o de mente manifiestamente más despierta, sin embargo, tomado todo en cuenta, las diferencias entre hombre y hombre no son tan considerables como para que alguien pretenda por ellas algún beneficio que otro no pueda pretender del mismo modo." (*Leviatán*, parte I, cap. XIII).

[94] Como es sabido, MacIntyre y sus seguidores reclaman la vuelta a una racionalidad teleológica del tipo de la descrita (excepto por la vinculación entre voluntad y acción, obviamente). Como señala Gauthier, se trata de modelos de racionalidad que ya incluyen un contenido moral (la referencia a fines). Entonces, lo que hay que probar no es la relación entre razón y moral, sino que, de hecho, la razón contenga tanto como estos teóricos afirman que contiene.

comunes más inmediatas sobre la racionalidad instrumental[95].

El desinterés mutuo es, como veíamos en el epígrafe anterior, uno de esos supuestos que, si bien no refleja necesariamente cómo somos de hecho las personas, sí puede ser aceptado metodológicamente como rasgo básico o previo, porque —aunque no seamos tan desinteresados— podemos reconocer reflexivamente que, si de lo que se trata es de seleccionar instituciones imparciales, preferimos dejar de lado nuestros afectos y comportarnos de modo tal que quede asegurado el mayor beneficio mutuo posible y la ausencia de explotación[96]. Generalizando, los supuestos metodológicos deben poder resistir el análisis reflexivo *ex post*, que se derivará del hecho de que las teorías desarrolladas a partir de tales hipótesis metodológicas respondan a la experiencia cotidiana y posean plausibilidad normativa (o, en palabras de Rawls, sean "razonables").

Pues bien, el contractualismo considera que el único modelo de individuo que resiste ese análisis reflexivo *ex post* se asemeja mucho al individuo económico. El desinterés mutuo es una de las características más sobresalientes de este individuo, pero no la única. A continuación examinaremos algunas de las restantes.

Para empezar, hay que decir que si por "individuo económico" entendemos el tipo de unidad individual que los economistas clásicos emplean para sus análisis, entonces los individuos en la posición inicial no son

---

[95] Trataremos pormenorizadamente en el punto siguiente los aspectos concernientes a la racionalidad de las partes. En este momento únicamente pretendemos dar algunas razones en favor de una caracterización económica de los individuos. Como quiera que es imposible deslindar completamente los temas, consideraremos este epígrafe conjuntamente con el punto siguiente, como formando una unidad. Si bien, queremos mantener la distinción entre el postulado individualista (aunque incluya una caracterización de las partes como agentes *económicamente racionales*) y la concepción de la racionalidad, justificada por razones argumentativas y filosóficas, no simplemente metodológicas.

[96] Estas discusiones nos alejan del tema que nos interesa aquí, pero sí conviene precisar —para evitar malos entendidos— que lo que realmente intentaría un individuo "mutuamente desinteresado" es explotar a los otros sin ser él mismo explotado. La ausencia de explotación se deriva de que este maquiavélico intento es, a su vez, mutuo.

exactamente individuos económicos, aunque posean algunos de sus rasgos típicos. Hecha esta salvedad, podemos comenzar por definir el *individuo económico* como un instrumento analítico, que no se refiere necesariamente a un ser humano individual, sino que suele remitir a la familia o la empresa (este es uno de los puntos en que el individuo contractualista se aparta del económico). Tal individuo es considerado racional "si y sólo si sus elecciones pueden ser representadas por una función de utilidad definida sobre todas las posibilidades alternativas de distribución de bienes y servicios (costes)"[97]. Si este es el caso, es decir, si es posible definir una función continua de utilidad para un individuo dado, entonces, conocidas sus circunstancias y las circunstancias del medio, es posible determinar (predecir) sus elecciones, o al menos (puesto que, en última instancia, el individuo es libre) explicarlas causalmente. Las condiciones de posibilidad de las funciones de utilidad, así como las condiciones externas que han de darse para que las elecciones de los agentes económicos puedan responder a sus funciones de utilidad, serán discutidas más adelante, al hablar de la racionalidad y el mercado.

Sí interesa destacar que el economista *no* define una función de utilidad para cada individuo y luego estudia cómo se comporta ese individuo respecto a dicha función (es decir, en qué medida trata de realizar su utilidad), sino que primero determina qué es lo que el individuo trata de alcanzar o realizar en el mayor grado (maximizar), y conforme a ello define sus utilidades. Esto quiere decir que el economista considera al individuo necesariamente como un *maximizador*.

La maximización es una característica analítica de los individuos para el economista clásico. Si un individuo sigue trabajando cuando ya tiene bastante para vivir, el economista dirá que trata de maximizar su seguridad mediante el ahorro; si deja de trabajar, dirá que trata de maximizar su bienestar mediante el disfrute de tiempo libre; si hace algo intermedio, dirá que trata de maximizar su preferencia, que consiste en una determinada combinación de bienestar y seguridad. Vemos, por tanto, que la relación entre utilidad y maximización no

[97] Gauthier, D., "Economic Rationality and Moral Constraints", *Midwest Studies in Philosophy*, III (1978), pp. 75-96; p. 76.

es contingente, sino necesaria: la utilidad se define según las preferencias, y éstas son aquello (sea lo que sea) que el agente satisface, luego se supone que el agente *siempre* intenta maximizar la utilidad, haga lo que haga. Así, hablar de individuo económico equivale a hablar de "maximizador de utilidad".

Pero la simple maximización de utilidad, que ciertamente nos informa sobre cómo actúan los agentes en interacción cuando carecen de interés por los intereses de otros (tratan de alcanzar en la mayor medida posible aquello que prefieren, sea lo que sea), necesita ser adjetivada para darnos alguna idea sobre los medios que poseen los agentes para alcanzar su objetivo maximizador y, sobre todo, las circunstancias limitadoras con que deben contar. Nos centraremos en dos adjetivos muy concretos que creemos definen bastante bien al sujeto económico. Como siempre, se trata de una reducción que dejará de lado algunos aspectos, pero confiamos que incluye los más relevantes. Estos adjetivos son la libertad de los agentes económicos y el hecho de que sean propietarios.

El neocontractualismo es un ejemplo de teoría moral liberal y no es sorprendente que herede no sólo la defensa del ideal ilustrado de autonomía e independencia individual, sino también la convicción de que la libertad de la voluntad es el componente básico de la individualidad. La tradición contractualista entiende la libertad en su sentido más elemental (libertad negativa), especialmente *en* el estado de naturaleza. Así, al inicio del capítulo XIV del *Leviatán*, Hobbes escribe

> "Por LIBERTAD se entiende, según el significado propio de la palabra, la ausencia de impedimentos externos."

Gauthier, por su parte, expresa la misma idea al referirse a la actividad libre de los agentes en el mercado:

> "...el presupuesto de *libre actividad* asegura que nadie está sujeto a ninguna forma de compulsión, ni a ningún tipo de limitación

que sus acciones no tuvieran ya en soledad."[98]

Como teoría liberal, el contractualismo defenderá una visión mucho más rica de la libertad: una visión que incluye la posibilidad de compromiso, aceptación de límites voluntarios a la acción, la coherencia intencional de los puntos de vista personales a lo largo del tiempo, la posibilidad de elegir lazos afectivos (afectividad libre), etc.[99]. Todo ello será descrito y discutido al final de la empresa de fundamentación moral, cuando ésta haya de prolongarse para alimentar un programa concreto de acción política y social (que habrá de incluir un catálogo de libertades civiles "positivas"), pero en este primer momento, basta con la caracterización simple de la libertad negativa. Los individuos en la posición inicial se conciben como seres libres de coacciones: eligen según sus preferencias sin restricción externa alguna. El único límite a sus decisiones está impuesto por las circunstancias (la escasez de ciertos bienes y su coste relativo). Estas circunstancias son fijas para el individuo pues, aunque puedan depender de las acciones de otros (en el sentido en que los precios dependen, en el mercado, de la demanda y la oferta), el número de variables es tan grande y la influencia de cada agente particular tan despreciable, que cabe concebir el resultado global de las acciones combinadas de todos como una situación estática y "objetiva"; como si se tratase de las condiciones físicas en una isla desierta. La ausencia de coacciones no significa, por tanto, que los individuos puedan conseguir todo aquello que desean o prefieren (pues existen límites de hecho), pero sí que nunca actuarán contra su voluntad a causa de la acción (o la mera presencia) de otros agentes[100].

---

[98] *MA*, p. 96.

[99] La extensa descripción de la concepción del individuo liberal puede verse en el capítulo XI de *MA*. En relación con él, puede verse la concepción de "individualismo ryleano" que R. Bhargava expone en *Individualism in Social Science* (*cit.*), p. 207 y ss., basada en la idea de que la acción y el conocimiento humanos son mucho más ricos de lo que aparecen según la Teoría Económica y el cognitivismo psicológico.

[100] Obviamente, no es una posibilidad descartada en realidad. De momento ha de tomarse como una hipótesis, más adelante se dará un fundamento racional para la misma, basado en la propia lógica de la teoría del contrato (Cfr. capítulo VII de *MA* "La posición inicial del negociación: los derechos y la salvaguardia").

Ahora bien, nos venimos refiriendo a circunstancias "objetivas" como si todas ellas fueran ajenas al propio sujeto; sin embargo, una de estas circunstancias está ligada al agente económico mismo: se trata de *su* propiedad, es decir, de aquel conjunto de capacidades y bienes de los que puede hacer uso exclusivo para satisfacer sus preferencias (sea consumiéndolos, sea transformándolos en otros bienes mediante el trabajo o el intercambio).

El concepto de propiedad empleado por los economistas suscita muy complejos problemas relacionados con el derecho (modos de adquisición originaria, legitimidad de los títulos, etc.); la justicia (justificación de la apropiación, aspectos de justicia distributiva) y la economía política (defensa del Estado liberal, entendido como asociación de poseedores; debate propiedad privada *vs.* propiedad estatal; el problema de los "bienes públicos", etc.). Lejos de tales controversias —sobre todo intentando quedar al margen, por el momento, del debate suscitado en torno al "individualismo *posesivo*"— el contractualista liberal adopta un concepto de propiedad exclusivamente en la medida en que es necesario para definir los contornos de la individualidad en un contexto de interacción que, por ser libre y mutuamente desinteresada, responde a los caracteres de una interacción típicamente económica[101].

Eliminados por hipótesis los lazos afectivos y los límites morales presentes en nuestras interacciones habituales, el estado de naturaleza que se va dibujando (ciertamente semejante al que describió Hobbes) reitera aquellos rasgos que los economistas clásicos atribuyeran al mercado perfectamente competitivo. Como es sabido, el mercado perfectamente competitivo (un ideal moral de libertad e igualdad para Adam Smith) se separa del estado de guerra hobbesiano en que supone —al margen de la ausencia de fuerza y fraude— la propiedad y el consumo privado de todos los bienes (y capacidades) que entran en el juego económico.

En el estado natural hobbesiano la individualidad se identifica exclusiva-

[101] La incomprensión de este enfoque liberal ha llevado a autores como Peter Danielson a suponer, erróneamente, que el agente-parte del contrato ha de estar necesariamente definido por referencia a sus *derechos de propiedad*, como el sujeto lockeano (Cfr. Danielson, P., "The Visible Hand of Morality", *Canadian Journal of Philosophy*, vol. 18, nº 2, Junio 1988, pp. 357-384; p. 369).

mente con la libertad, y entonces "se seguiría que en tal condición, cada hombre tiene Derecho a todas las cosas; incluso al cuerpo de los otros"[102]. Allí donde la libertad se lleva a sus últimas consecuencias, la propiedad es precaria (aunque ello no quiere decir que no exista; al menos como demanda, o derecho). Tal precariedad *excluye* una interacción económica racional[103]. Sólo en la medida en que podemos reconstruir racionalmente el proceso mediante el que cada individuo, en uso de su razón y en vista de su interés (la auto-conservación), decide voluntariamente entrar en el pacto social y convenir con sus semejantes el respeto a la propiedad; sólo en esa medida, decíamos, podríamos considerar la propiedad como una de las características del individuo hobbesiano. El individuo hobbesiano puede ser calificado con razón como *homo oeconomicus*, pues la racionalidad pre-moral que exhibe *es* una racionalidad económica. Pero el tipo de interacción natural en la que participa no es económica porque le falta el elemento estabilizador del respeto a la propiedad privada.

Como sabemos, Locke supo ver esta necesidad, y concibió un estado natural moderado por ciertos límites morales. La presencia de dichos límites (en forma de derechos naturales) erradica la posibilidad de una moral por acuerdo, pero, paradójicamente, habilita el camino hacia un pacto social que no dependa, para su sustento, de la amenaza de Leviatán.

El contractualista moral necesita tener abierto el camino que transitó Locke, pero no puede aceptar la presencia de derechos previos en el estado de naturaleza, porque ello echa por tierra su compromiso de justificación exclusivamente racional de todas las restricciones morales (y los derechos producen "restricciones morales indirectas"[104]). Así podemos entender la necesidad de concebir al individuo como un ser capaz de interactuar con plena

---

[102] T. Hobbes, *Leviathan*, part. I cap. XIV.

[103] Al menos en principio. Recientes lecturas de Hobbes, a las que aludiremos en el próximo capítulo, exploran la posibilidad de que cierta interacción "racional" surgiera incluso en las condiciones naturales descritas por Hobbes.

[104] El término entrecomillado procede del gran teórico contemporáneo de los derechos (en buena medida sucesor de Locke, por cierto), Robert Nozick. Cfr. *Anarquía, Estado y Utopía*, México, Fondo de Cultura Económica, 1988, p. 42.

racionalidad económica con sus semejantes —y, por tanto, como propietario— pero, al mismo tiempo, la imposibilidad de afirmar que la propiedad está basada en derechos naturales o en estados anteriores. Por lo tanto, la propiedad (y los títulos en que se funda) posee el mismo estatuto hipotético que los restantes rasgos que definen la individualidad. Ha de esperar a la conclusión de la teoría para recibir una justificación definitiva.

En este punto, la propiedad juega el papel de servir a la construcción de una individualidad (hipotética) de mercado. Más abajo[105] debatiremos en conjunto los caracteres del mercado y su función en la teoría de Gauthier, así como los problemas que suscita la justicia en la distribución inicial de factores (propiedad). De momento, baste decir que la cantidad de bienes y factores de producción que un individuo posee (que llamaremos en adelante su "dotación") unidos a sus deseos o preferencias, sirven para definir exactamente su función de utilidad *y* el nivel de satisfacción que puede alcanzar. Es decir, dada una dotación *x*, *existe* un máximo de contrapartidas que el individuo puede obtener del mercado, y será una decisión personal qué distribución concreta de bienes o servicios representa, para él, una utilidad mayor.

Gauthier argumenta que todos los individuos se identifican con su función de utilidad (pues ésta expresa lisa y llanamente sus preferencias). Si, además, un individuo se identifica con su dotación inicial, entonces podemos decir que ese individuo se identifica con su "yo de mercado". El "yo de mercado" de una persona queda definido por su función de utilidad y su dotación inicial. Entre ambos (función de utilidad y dotación) fijan las preferencias y capacidades de la persona, que son los factores relevantes para su actividad en el mercado[106].

El "yo de mercado" es un concepto muy restringido a un tipo de interacción concreta. Toma los conceptos de libertad, propiedad y racionalidad maximizadora —así como la interacción libre en un marco artificial inexistente, como es el mercado idealmente competitivo— en su sentido más radical.

Se trata de un concepto-límite, que el contractualista moral expone para

---

[105] Cfr. punto 4, en este mismo capítulo.

[106] Cfr. Gauthier D. *MA*, p. 86.

evidenciar hasta qué punto es posible imaginar un modelo de individualidad abstracta y, sin embargo, receptora de los caracteres esenciales que definen el esquema de la mayor parte de nuestras interacciones[107]. El contractualista moral no necesita, en puridad, partir de un concepto semejante de individualidad. De hecho, Gauthier recalca las diferencias entre el individuo contractualista y el económico en su artículo "Economic Rationality and Moral Constraints"; veámos algunas de ellas:

En primer lugar, la familia o la empresa, unidades del análisis económico, *no* son la unidad apropiada para el análisis moral[108]. La unidad del análisis moral es la persona singular. Ello es válido tanto para las teorías que pretenden asignar derechos naturales a las personas (Locke, Nozick), como para quienes consideran que el respeto a ciertos derechos ha de derivarse de una situación anterior en la que no estaban presentes como tales (Hobbes, Gauthier, Rawls). Esta diferencia es crucial porque la familia o la empresa pueden tomarse como individuos continuos (no perecen), cuya propiedad o derechos —que se transfiere mediante intercambio— puede tomarse como algo dado una vez para siempre, de modo que se evitan cuestiones sobre la justicia de la distribución. Sin embargo, si la unidad del análisis es la persona física, entonces surgen continuamente cuestiones sobre la legitimidad de la distribución de propiedades y derechos, ya que continuamente hay miembros nuevos que *entran* en la sociedad, y otros que salen. Este problema ocupa, por ejemplo, una de las partes centrales de la obra de Nozick, pues es difícil conciliar la idea del origen *tradicional* de los títulos con la intuición de que cada nuevo miembro de la sociedad viene ya provisto de ciertos derechos inalienables. También Rawls (y, en general, todas las teorías contractualistas) tiene que vérselas con

---

[107] Mas no se identifica totalmente con el agente racional que fungirá como parte del contrato. Entre otras diferencias, el "individuo económico" está provisto de derechos previos a la interacción de mercado; el agente racional-parte del contrato posee, a lo sumo, capacidades, pero ningún derecho previo. La insistencia en este punto se debe a que la equivocada creencia de que el contractualismo de Gauthier exige suponer derehos anteriores al pacto ha provocado una absurda crítica por parte de P. Danielson (Cfr. más arriba, nota 101, p. 100). La confusión de Danielson debe aclararse definitivamente más abajo, al exponer sistemáticamente la teoría.

[108] Cfr. Gauthier, D., "Economic Rationality and Moral Constraints" (*cit.*), p. 84.

el problema del asentimiento al pacto de las generaciones futuras. En la *Teoría de la Justicia* Rawls lo resuelve suponiendo que los sujetos en la posición originaria *representan* líneas familiares, solución deudora del análisis económico. Gauthier tratará de no caer en esta "ontología económica". Él mantiene que los sujetos en la situación inicial *son* personas individuales, que sólo se representan a sí mismos. La fuerte caracterización como individuos auto-interesados e instrumentalmente racionales le va a permitir sostener que cualquier miembro de la sociedad puede figurarse *ex post* la posición inicial y reconocer que, *aun en el peor de los casos y conociendo (ahora) plenamente sus circunstancias*, habría asentido al pacto suscrito. Esta explicación justificará el mantenimiento del pacto por las generaciones posteriores.

Por último, hemos de insistir en un punto que mencionábamos arriba y que es una de las claves para distinguir el individuo hobbesiano del individuo tal como es concebido por el contractualista contemporáneo. El economista persigue simplemente, como hemos visto, una estrategia analítica para explicar las decisiones de los agentes económicos y los resultados de sus interacciones. Su instrumento analítico resulta ser un conjunto de individuos poseedores cuyos títulos sobre aquello que poseen no se discuten (tal discusión excede el análisis económico). Además, estos individuos pueden no ser personas físicas, sino otro tipo de entidades (cuya inadecuación para el análisis moral es evidente). Pues bien, a estas dos características que hemos expuesto, se añade la concepción de los individuos como meras "máquinas de maximizar", es decir, como agentes cuya racionalidad termina en el cálculo utilitarista inmediato *de* cada situación de interacción concreta. Por así decir, el agente económico carece de una visión coherente a largo plazo. Aunque pueda hacer *ahora* aquello que cree que le procurará mayor beneficio a largo plazo, puede (y debe, desde su punto de vista) hacer *más tarde* lo opuesto si entonces calcula que es eso lo más beneficioso. La relación de su acción con el beneficio es tan íntima, que excluye prácticamente toda posibilidad de compromiso para el futuro.

El concebir al individuo de un modo tan recortado da lugar a los conocidos dilemas de la racionalidad (tipo Dilema del Prisionero) cuyo antecedente más respetable es, tal vez, el razonamiento del Tonto (*Foole*) que

nos presenta Hobbes. Éste representa la persona que, por un lado, ve que es beneficioso para ella suscribir un pacto, pero luego, viendo que es aún más beneficioso incumplirlo, no es capaz de darse razones suficientes para el cumplimiento. Es decir, queda atrapada por un tipo unidimensional de razonamiento maximizador. Hobbes supuso que *somos* así, y que el único modo de mantener los pactos es hacer que defraudar sea aún más costoso que cumplirlos (mediante la incorporación al mismo del apropiado aparato represivo). Gauthier y los contractualistas liberales contemporáneos no están de acuerdo con Hobbes. El individuo se concibe, según hemos visto, como un ser tan egoísta y dominado por sus deseos e intereses particulares como lo concibiera Hobbes, pero también se admite que *es posible* que el mismo afán maximizador conduzca a ciertos sujetos a reconocer y adoptar nuevos modos de interacción mutuamente beneficiosos. En esto consiste la fuerza del neo-contractualismo. De hecho, es uno de sus nudos teóricos, que debatiremos más tarde. Aquí destacamos que, a pesar de la caracterización económica de los sujetos, éstos no son concebidos por el contractualista como "máquinas de maximizar", sino como seres capaces de reflexionar sobre su modo de razonar y, eventualmente, adoptar nuevos principios para la acción sobre la base de esa reflexión.

Las notas que hemos dado hasta ahora definen lo principal de la concepción metodológica de individuo adoptada por Gauthier (representante del "contractualismo moral liberal"). Ha quedado claro también el papel de este postulado en la teoría contractualista: caracteriza a las partes en la situación originaria, de modo que sea factible reproducir heurísticamente el acuerdo hipotético a que tales partes llegarían. En estas últimas páginas hemos dado, creo, una sensación errónea, porque hemos tratado el neo-contractualismo como si compartiera cierta concepción del individuo (opuesta, acaso, al individualismo hobbesiano o lockeano). Nada más alejado de la realidad. De hecho, como señalábamos al inicio del capítulo, la teoría del contrato social tiene un funcionamiento tan diáfano que si se parte de premisas iguales se *deben* alcanzar resultados iguales. Como quiera que los resultados no son iguales, ello indica que, probablemente, se ha partido de premisas ligeramente diferentes.

Si dos contractualistas como Rawls y Gauthier llegan a resultados tan distintos, ello se debe a que adoptan puntos de partida distintos. Nosotros hemos inclinado nuestra descripción del postulado individualista del lado de Gauthier (aunque no sistemáticamente) y, precisamente por eso, creemos conveniente ahora decir algo sobre su referente polémico, que es la concepción de las partes en la posición originaria de Rawls.

g) Excurso: el referente rawlsiano.-

Una porción nada despreciable de la caracterización de las personas en la posición inicial se debe a la insatisfacción que Gauthier experimenta respecto a la descripción de las partes hecha por Rawls. Gauthier siente que Rawls no logra superar la crítica que él mismo formuló al utilitarismo (del cual dice que no toma en serio la distinción entre las personas). Su artículo "Justice and Natural Endowment: Toward a Critique of Rawls's Ideological Framework" —de gran importancia para la fijación de la postura teórica de Gauthier, como afirmábamos en el capítulo anterior— recoge, ya en 1974, los argumentos en contra de la caracterización rawlsiana de las partes, frente a la cual, los rasgos del individuo tal como es concebido metodológicamente por el contractualismo liberal aparecen con más claridad[109].

[109] La crítica gautheriana a Rawls podría haberse inspirado en (o, en todo caso, coincide con) las reflexiones de Thomas Nagel en *The Possibility of Altruism* (Oxford, Clarendon, 1970). Éste (ciertamente, sin pensar en la teoría de Rawls) escribe: "Para conceder el apropiado peso a las necesidades, deseos e intereses de todos los individuos, se debe requerir que la elección de un principio interpersonal de valoración se realice bajo la condición de que el decisor espere tratar todas las vidas en cuestión, no como una sóla super-vida, sino como un conjunto de vidas individuales diferentes, cada una de las cuales fuese un conjunto completo de experiencias y actividades. Si tal procedimiento de decisión pudiera hacerse inteligible, garantizaría ciertamente la demanda individual de poseer una voz igual en la consideración de qué principio de valoración adoptar -una voz que se le otorgase no como una vida posible, sino como una vida real. Pero no está claro cómo podría hacerse esto" (p. 141). Gauthier coincide plenamente con el diagnóstico de Nagel y, por así decir, acepta el reto. No obstante hay que precisar que, probablemente, en 1970 Nagel pensaba que la Teoría de la Justicia de Rawls (por entonces formulada en "Justice as Fairnes" {*The Journal of Philosophy*, 54 (1957), pp. 653-670} y "Distributive Justice" {en P. Laslett y W.G. Runciman, *Philosophy, Politics and Society*, Londres, Blackwell, 1967, pp. 58-82},

El núcleo de la crítica contractualista-liberal a la concepción rawlsiana de las partes podría expresarse con una metáfora asimismo rawlsiana: el "velo de ignorancia" tendido por Rawls es demasiado denso, innecesariamente espeso. El grosor excesivo del velo es causa de algunas de las características de la concepción de la justicia inaceptables desde un punto de vista liberal. La fuerza de esta crítica reside en que nunca niega que el marco ideológico de Rawls *es liberal*[110]. Desde esta óptica, el innecesario grosor del velo de ignorancia representa una incorrecta apreciación de las consecuencias normativas de la aceptación de la racionalidad instrumental de las partes y el individualismo metodológico. La tesis de Gauthier es sencilla: veamos qué ocurre si tomamos en serio los presupuestos (liberales) de Rawls. Y lo que ocurre es, según él, que la posición original no cumple uno de los requisitos que el propio Rawls establece para ella: el que sea interpretada de modo que en cualquier momento pueda adoptarse su perspectiva. Para que la posición original se adecue a este requerimiento, ha de ser reformulada de forma que, eliminando algunas capas del velo de ignorancia, el individuo aparezca menos borroso, más distinguible por sus intereses y capacidades.

El contractualismo liberal se compromete a demostrar que, sobre esta nueva concepción de las partes (y sólo sobre ella), es posible reconstruir un pacto normativo —logrado tras una negociación que ejemplifica el tipo de interacción pre-moral entre sujetos racionales— que *sí* resiste el *test* post-acuerdo que no soportan los principios de la justicia rawlsianos. Algunas de las consecuencias de este compromiso fueron ya comentadas en el capítulo anterior[111], y sus frutos definitivos serán expuestos y debatidos al comentar la teoría de Gauthier. En este momento quisiéramos centrarnos exclusivamente

---

aunque seguramente conocida en su versión de 1971 por Nagel, que trabajaba en Harvard junto a Rawls) sería capaz de superar la limitación del utilitarismo. Es mérito de Gauthier el haber visto, ya en 1974, que la teoría de Rawls *tampoco* tomaba en serio la separación entre presonas y el haber intentado superar con su teoría esta carencia.

[110] Gauthier da por sentada "La suposición por parte de Rawls de esa concepción de la razón que es prevalente en (y fundamental para) nuestra sociedad caracteriza su propio *marco ideológico* y lo identifica, en un aspecto esencial, con lo que denominaré neutralmente el marco *individualista liberal*." ("Justice and Natural Endowment...", *cit.*, p. 152.

[111] Cfr. cap. 1, punto 5.

en la diferente concepción de las partes, sus razones y sus consecuencias inmediatas. Esperamos mostrar que esta diferente concepción presenta de modo privilegiado la divergencia entre el contractualismo liberal y el contractualismo de Rawls, con lo que ayuda a definir los contornos del primero.

Comenzaremos con una pregunta: ¿Por qué considera Gauthier que Rawls sigue sin tomar en serio la distinción entre las personas? La respuesta está contenida en el siguiente párrafo:

> "Tras el velo de ignorancia, las personas están idénticamente situadas, no sólo en sus circunstancias objetivas, sino también subjetivamente, puesto que cada uno es completamente ignorante de sus capacidades e intereses y, por tanto, incapaz de distinguirse a sí mismo de sus semejantes."[112]

La confusión entre las personas es incompatible con una consideración seria de la individualidad y, además, elimina la posibilidad de un verdadero contrato, ya que las partes "no tienen ninguna base para negociar unas con otras, con lo que el acuerdo sobre los principios de la justicia puede ser representado por la decisión de un solo individuo representativo"[113]. Ello desvirtúa la naturaleza propia del contractualismo —de hecho, es una de las características de la teoría rawlsiana que más la acerca a Kant y la separa de Hobbes, Locke e incluso de Rousseau. Sin embargo, esta reducción del papel de un verdadero pacto en la teoría de la justicia de Rawls, con ser contraria a la tradición contractualista liberal, no es la principal fuente de insatisfacción

---

[112] Gauthier, D., "The Incompleat Egoist", *cit.*, p. 237. La idea de este párrafo se basa en un famoso texto que podemos leer en la p. 163 de la versión española de la *Teoría de la justicia*: "nadie conoce su lugar en la sociedad, su posición o clase social; tampoco sabe cuál será su suerte en la distribución de talentos y capacidades naturales, su inteligencia y su fuerza, etc. Igualmente nadie conoce su propia concepción del bien, ni los detalles de su plan racional de vida, ni siquiera los rasgos particulares de su propia psicología...". Gauthier resume este texto diciendo que, según Rawls, "*nadie sabe quién es*" tras el velo de ignorancia (Cfr. "Justice and Natural Endowment..." *cit.*, p. 155).

[113] Gauthier, D., "The Incompleat Egoist", *cit.*, p. 237.

desde el punto de vista de Gauthier. Porque, en la medida en que el individualismo radical de estas teorías es puramente metodológico, la concepción abstracta del individuo rawlsiano se podría aceptar si los resultados de su teoría fuesen plausibles o razonables. Y el hecho es que, según Gauthier, no lo son.

La plausibilidad del argumento contractualista se verifica si, desde la situación particular de cada persona en la sociedad, es posible adoptar de nuevo la posición original y asentir al razonamiento (o discusión) que condujo a la adopción de los principios de la justicia. Dicho de otra forma, el argumento contractualista resultará plausible desde el punto de vista particular de cada miembro de la sociedad si y sólo si cada uno puede comprobar que, dada su situación, ningún otro principio de distribución de los beneficios sociales *aceptable por todos* le habría ofrecido mejores perspectivas. Rawls sostiene que su argumento cumple esta condición de plausibilidad. Gauthier lo niega.

Para justificar su postura, Gauthier debe introducir un concepto que Rawls "olvida" (al menos en parte[114]). Se trata de la noción lockeana —pero de ascendencia hobbesiana— de "dotación natural": la idea de que cada individuo posee algo por naturaleza. Es, sin duda, cuestionable cuál sea el contenido, si alguno, de la "dotación natural". Como sabemos, Locke extiende esta dotación a la libertad, las capacidades, los bienes logrados con el propio esfuerzo o trabajo y los apropiados para el uso privado, siempre que "queden suficientes bienes comunes para los demás"[115]; pero esta concepción es sólo un ejemplo. Según nuestra interpretación, el sentido último de la idea de "dotación natural" y, por tanto, aquello a lo que el liberalismo no puede renunciar (por encima de las diversas interpretaciones), es el hecho de que los hombres son *distintos* en sus capacidades, ambiciones, deseos e intereses. Tal

---

[114] En las pp. 95-96 de la *Teoría de la justicia* leemos: "No hay mejor razón para permitir que la distribución del ingreso y la riqueza sea resuelta en función de las capacidades naturales, a que lo sea en función de las contingencias sociales e históricas". Este texto resulta paradigmático del modo en que Rawls trata las capacidades naturales (lo que llamaremos "dotación natural"): no niega que existan, pero las considera arbitrarias desde un punto de vista moral, por lo que no cree que deban ser tenidas en cuenta al elegir criterios de justicia distributiva. Al contrario, los principios de justicia deben "mitigar los efectos arbitrarios de la lotería natural" (*ibid.*).

[115] Cfr. Locke, J., *Segundo tratado sobre el gobierno civil*, Madrid, Alianza, 1990, cap. 5 (en esp. parágrafos 25-32). La cita procede de la p. 57.

distinción tal vez no aparezca muy evidente en un estado de naturaleza como el descrito por Hobbes, donde ni siquiera los más capacitados pueden esperar beneficio alguno (sobre todo en relación con un hipotético estado social). Pero la distinción entre los hombres, y la "dotación natural" que es su origen, siguen existiendo: para el liberal son una evidencia que no se puede negar sin negar el fundamento de nuestra individualidad. Rawls no niega esta evidencia pero, al reclamar que los principios de la justicia "mitiguen" los efectos de la lotería natural, niega que las partes posean derecho alguno sobre las capacidades y energías personales que les individualizan. Rawls considera, por tanto, que las capacidades naturales son comunes; todos (hayan sido naturalmente favorecidos o no) tienen el mismo derecho a beneficiarse de ellas. Pero como las capacidades naturales, a diferencia de los bienes, no son separables de las personas que las disfrutan, su distribución crea problemas que Rawls deja sin solucionar[116].

Gauthier inicia su crítica en ese punto, sobre la base de la evidencia de que cada persona posee preferencias y capacidades diferentes. Sobre tal base, y aun sin discutir cuál sea el contenido de la dotación natural, cabe postular un contexto de interacción no-social (o punto de no-acuerdo) que arroja cierto resultado (beneficio) para cada individuo. El resultado no será en absoluto arbitrario. Tendrá que ver, obviamente, con la dotación de cada individuo, que comprende con seguridad, al menos, sus capacidades y los intereses en cuya virtud las orienta. La interacción social o cooperativa que surge tras el acuerdo sobre los principios de distribución justa puede compararse, así, con un modelo de interacción en el que la distribución del beneficio es "natural"[117]. No se afirma que tal tipo de interacción pre-social sea posible, pero sí que es pensable, pues no contradice las experiencias inmediatas o reflexivas sobre el

[116] El más grave de ellos es, sin duda, el problema del asentimiento *ex post* de un individuo que, habiéndo sido muy favorecido por la "lotería natural", se encuentre, tras el pacto, en una posición de contribuyente neto a la empresa social. En torno a esta dificultad girará la crítica de Gauthier.

[117] La hipotética distribución resultante según este modelo servirá como "posición original", "situación inicial" o "*base-line*".

sentido y contenido de la individualidad (más bien es apoyado por ellas)[118]. En la medida en que cabe pensar un tipo natural de interacción, y dada la conciencia individual de poseer cierta dotación natural, cada individuo puede, una vez levantado el velo de ignorancia y conocida cuál es su dotación natural real, verificar la plausibilidad del argumento contractualista *comparando* la utilidad que le corresponde según los principios de la justicia con aquella que de todas formas habría obtenido en una situación de no-acuerdo. En palabras de Gauthier:

> "Cada ser humano es un agente con ciertas preferencias y ciertas capacidades físicas y mentales que, en ausencia de otros, orienta naturalmente a la satisfacción de sus preferencias. Esto proporciona una base, de ningún modo arbitraria, desde la que podemos examinar y valorar la interacción, introduciendo concepciones tales como 'mejorar' y 'empeorar'. Un principio que haga abstracción de esta base no tendría en cuenta que los seres humanos son agentes. Un principio que no considerara esta base como normativamente fundamental, no se dirigiría imparcialmente a los seres humanos como agentes"[119]

Pues bien, Gauthier sostiene que los principios de la justicia de Rawls son deducidos haciendo abstracción de esa base, por lo que no resultarían plausibles para agentes racionales tras levantar el velo de ignorancia. La causa de esta implausibilidad estriba en que cada individuo es capaz de realizar una distinción entre beneficios procedentes de la sociedad (a la que Rawls define, recordemos, como una "empresa cooperativa encaminada al beneficio mutuo")

---

[118] Gauthier lo explica, en "Justice and Natural Endowment" (*cit.*,), p. 159, con las siguientes palabras: "Tras el velo de ignorancia nadie conoce sus capacidades y talentos naturales y, por lo tanto, nadie sabe qué podría obtener en ausencia de acuerdo. Sin embargo, cada uno sabe que posee ciertas capacidades y talentos naturales, y que la gente difiere en esta dotación, de modo que, incluso en ausencia de acuerdo, la gente tendría asegurados diferentes niveles de bienestar. Por tanto, es posible que cada uno tome en cuenta el 'punto de no-acuerdo' en su razonamiento, aunque ninguna persona particular sepa cómo le afectaría a ella."

[119] Gauthier, D., *MA*, p. 221.

y aquellos otros bienes que cada uno podría alcanzar o producir en cualesquiera circunstancias. En la medida en que se acepta que *todos* poseen una dotación natural distintiva, cualquiera puede calcular qué parte de los bienes que actualmente disfruta es un verdadero "beneficio cooperativo" y qué parte corresponde a lo que de todas formas habría obtenido en caso de no-acuerdo. La primera consecuencia de este contraste es evidente: una parte de los bienes que las personas disfrutan *no* pueden considerarse producto de la empresa cooperativa que llamamos sociedad, sino que habrían sido igualmente producidos por las capacidades y talentos de cada individuo aplicados únicamente al objetivo de maximizar sus intereses particulares, sin planear actividad cooperativa alguna. De ahí se sigue que, a la hora de acordar un principio conforme al cual distribuir el beneficio que se espera alcanzar con la cooperación, haya que discriminar entre este beneficio y la parte de bienes no-cooperativos que, aunque se obtendrán *en* la sociedad, no proceden *de* ella. Si el principio de justicia tiene en cuenta esta distinción, entonces cualquier persona (concebida como agente) puede identificarse con el pacto originario, ya que se verá a sí misma como propietaria, en primer lugar, de su "dotación natural" y lo que podemos llamar "rendimientos" de la misma y, en segundo lugar, de una parte de los beneficios cooperativos (la parte que le corresponda según el principio de distribución justa acordado[120]).

El principio de la diferencia, debido a las condiciones de extrema incertidumbre bajo las que es elegido, no permite que las personas reales se identifiquen retrospectivamente con el pacto originario, ya que no discrimina

---

[120] Es importante hacer aquí dos advertencias. La primera es que el argumento de Gauthier incluye la defensa de un principio de distribución alternativo al segundo principio de la justicia de Rawls (principio de la diferencia); sin embargo nosotros intentamos ceñirnos a la parte del argumento que tiene influencia en la concepción de la partes, por lo que no debatimos ese principio alternativo en este lugar. La segunda se refiere al uso que estamos haciendo del concepto "dotación natural". Tal vez dé la sensación, en algún punto, que dicho concepto supone la defensa de "derechos naturales". Es una sensación explicable, porque la idea de una "dotación natural" está conectada (aunque no directamente) con la existencia de derechos naturales. Sin embargo, ése es otro aspecto que debemos dejar de lado de momento, pues el modo preciso en que esta conexión se realiza sólo aparecerá claro tras la justificación racional de una moral por acuerdo. En ese momento podremos discutir versiones alternativas (Nozick, Dworkin). Por ahora, aceptaremos que la dotación natural no tiene, en principio, implicaciones morales directas. La pretensión que un agente tiene sobre su propia dotación natural y los rendimientos obtenidos en un estado no-social no será más que una "demanda racional"; aún no una reclamación moral.

entre los dos tipos de bienes que hemos distinguido. El individuo representativo que elige tras el velo de ignorancia "olvida" que cada individuo concreto posee unas capacidades, talentos y preferencias específicos que deberían reflejarse en su elección. Este individuo representativo elige un principio para distribuir equitativamente *todos* los bienes sociales (sin distinguir si son frutos de la cooperación o no). Con ello toma las capacidades y talentos naturales como si fueran compartidos, como si fueran un rendimiento más de la empresa cooperativa. Sin embargo —siguiendo con el símil económico— tales talentos y capacidades no son rendimientos de la empresa social sino, más bien, su capital inicial, aquél que es debido a cada uno de los accionistas antes de repartir (equitativamente, por supuesto) los beneficios. Y ha de notarse que esa "devolución" a cada accionista de su aportación inicial a la empresa no prejuzga el modo en que se distribuirán los beneficios. Esta distribución podría incluso ser "solidaria" o "benéfica" (de modo que se retribuyera más a quien aportó menos) si así se decidiera. Pero lo que nos parecería completamente injustificado es que, con el afán de equilibrar la distribución final de bienes, alguna de las partes, no sólo se quedase sin participación alguna en los beneficios, sino que además hubiera de ceder una porción de lo que originalmente aportó, con lo que quedaría convertida en contribuyente neta a la empresa.

Gauthier se pregunta qué incentivo tendría para participar en la sociedad un individuo que no sólo no obtiene beneficio alguno de la misma, sino que además incurre en un coste que podría evitar permaneciendo fuera. Esta situación (tal vez difícil de imaginar, pues cabe suponer que los beneficios ligados a la sociedad exceden cualquier posible coste) no es imposible, dados los presupuestos rawlsianos y la formulación de su segundo principio. Y si es posible, entonces los principios de la justicia resultarán implausibles para algunos de los individuos[121], y la posición originaria habrá fracasado en su objetivo de ser un ideal normativo cuya perspectiva puede ser adoptada en

[121] En principio, para aquellos que se perciban a sí mismos como contribuyentes netos a la empresa social; pero no sólo. Todo individuo racional (aunque obtenga "cierto" beneficio de la cooperación social) puede comprobar si ese beneficio es *tanto como podría haber obtenido* si la cooperación estuviera basada en un principio de distribución alternativo unánimemente aceptable. Puesto que los individuos se conciben metodológicamente como auto-interesados, no aceptarán el principio propuesto por Rawls si ese otro principio alternativo existe.

cualquier momento por cualquier sujeto. El argumento de Rawls no tiene en cuenta qué tipo de motivación poseen los individuos *antes* de entrar en la sociedad. Entre individuos mutuamente desinteresados (como él mismo los define) sólo la expectativa de cierto beneficio conducirá lógicamente a un contrato social. Se hace así evidente que existe al menos un principio equitativo de distribución alternativo *que habría sido aceptado por todos* en una situación original: aquél que tenga en cuenta la distinción entre "dotación natural" —aquella parte de bienes que se deben inexcusablemente a cada miembro de la sociedad, como restitución de su aportación a la empresa cooperativa— y beneficio cooperativo —aquella otra parte que ha de ser equitativamente dividida entre todos, por ser un producto de la cooperación social que no habría podido ser alcanzado por cada uno en solitario. Sólo un principio así incorporaría el incentivo necesario para que la empresa social motive suficientemente a *todos* los agentes. No discutiremos aquí en detalle la base racional del principio alternativo propuesto por Gauthier. Simplemente resaltamos que el hecho de tener en cuenta la dotación natural de cada individuo y *retribuir* a cada uno con relación a la misma (sin perjuicio de la distribución más o menos igualitaria del beneficio cooperativo) asegura que, por así decir, nadie podrá denunciar el contrato social por los perjuicios que le causa; nadie se verá a sí mismo como contribuyente neto a la empresa social ni verá a ningún otro miembro de la misma como un "aprovechado". Al contrario, *cualquier* individuo tendrá un motivo racional (fundado en su propio beneficio) para participar en la empresa social y cumplir el pacto originario. Como ha escrito David Braybrooke:

> "En principio Gauthier ha resuelto, hasta donde un argumento racional puede resolverlo, el problema del asentimiento que persistía en el argumento de Rawls para los agentes que, una vez levantado el velo, encontraban que no pertenecían al estrato de los menos-aventajados. Todos los agentes de Gauthier, ganadores o perdedores en su carácter y capacidades tienen fundamentos que apelan a su individualidad para aceptar la estructura que la justicia prescribe, porque cada uno de ellos estuvo representado,

individualmente, en su elección."[122]

Desde otra perspectiva, el fracaso de los principios de la justicia desde el punto de vista liberal puede explicarse diciendo que, al considerar que la distribución natural de talentos, capacidades y deseos es "arbitraria desde el punto de vista moral", Rawls elimina cualquier posible fundamento para la identidad personal. Según Gauthier, Rawls confunde el hecho de que tras el velo de ignorancia nadie es consciente de *su* identidad (es decir, de *quién* es él o ella), con el hecho de que nadie sea consciente de que posee *alguna* identidad:

> "Porque aunque el agente ideal no es consciente de su identidad, es consciente de que tiene una identidad. Parece razonable, entonces, que elige con esto en mente, y considera su reclamación sobre los frutos de la interacción social dada esa identidad. En ese caso, elegiría un principio que regulase la interacción de tal modo que las características naturales particulares de cada persona, en tanto afecten a lo que realiza, se tuvieran en cuenta en la determinación de la distribución de los beneficios. (...) Debemos suponer que las características de cada persona, que le capacitan para realizar cierta contribución al producto social, le dan también derecho a cierta porción de ese producto."[123]

En conclusión, esta reflexión retrospectiva propuesta por Gauthier descalifica la deducción rawlsiana de los principios y cuestiona la caracterización de la posición original desde la que se realiza. Gauthier considera que el error de Rawls no está tanto en su primera descripción de las partes (individuos racionales, mutuamente desinteresados, etc.) sino en haber aplicado, como garantía moral, un velo de ignorancia que oscurecía innecesariamente las

---

[122] Braybrooke, D., "social Contract Theory's Fanciest Flight", *Ethics*, 97 (julio 1987), pp. 750-764; p. 754.

[123] Gauthier, D., *MA*, pp. 251-252.

diferencias entre individuos hasta hacerlas desaparecer, eliminando así la posibilidad de un verdadero contrato que reflejase los intereses de cada uno y que cada uno pudiese reproducir retrospectivamente como criterio normativo ideal. Rawls creyó que el punto de vista moral sólo podía ponerse en relación con el contractualismo si se postulaba un "estado de naturaleza" en condiciones de gran incertidumbre[124].

Al desarrollar una teoría moral contractual liberal (alimentada, en buena medida, por el debate que hemos resumido), Gauthier espera haber hallado el modo de superar el dilema planteado por su crítica a Rawls. Los cuernos del dilema contractualista se podrían enunciar así: primero, es posible ofrecer una base contractual de la justicia, pero sólo si se oscurecen los presupuestos liberales del contrato; segundo, si se respeta la coherencia de las premisas liberales, es posible legitimar contractualmente las obligaciones políticas, pero no las obligaciones morales —como mucho, se produce únicamente la justificación de una moral mínima, reducida a la eliminación (coactiva) de la fuerza y el fraude del mercado[125].

Gauthier sostiene que es posible encontrar un fundamento contractual para la moral *sin* renunciar a los presupuestos liberales; sin, por ejemplo, tener que ocultar las diferencias individuales tras un velo de ignorancia. Esta tesis es expresada de forma contundente en *MA*:

> "La identidad de una persona es, desde cualquier perspectiva, un asunto contingente. Pero esta contingencia no es moralmente

[124] En honor a la verdad hay que decir que el Gauthier de 1974, cuando escribe "Justice and Natural Endowment...", estaría de acuerdo con esta apreciación de Rawls. De hecho, la conclusión de su crítica a la configuración de la posición original es que partir de presupuestos liberales (y mantener la coherencia) *impide* alcanzar conclusiones normativas que puedan considerarse morales. Así, lo que criticaba entonces Gauthier es que Rawls, en tanto pretendía haber alcanzado conclusiones normativas universales, había sido incoherente con los principios que inspiran su teoría. Literalmente, Rawls habría corrido un tupido velo sobre sus principios liberales para poder ofrecer conclusiones acordes con su propio sentido (o visión intuitiva) de la justicia. Gauthier simplemente hace ver cuáles eran aquellos principios liberales y cómo Rawls los obvió.

[125] La idea de un "dilema contractualista" (que, en términos generales, podría identificarse con la formulación que he propuesto) procede de Jung Soon Park, *Contractarian Liberal Ethics and the Theory of Rational Choice*, Nueva York, Peter Lang, 1992, cap. III, en especial pp. 94 y ss.

> arbitraria, porque *la moralidad se halla y sólo puede hallarse en la interacción de personas reales, individualizadas por sus capacidades, actitudes y preferencias. En nuestro análisis nos tomamos en serio la individualidad de las personas.*"[126]

Así, en *MA*, los argumentos de una crítica liberal al "igualitarista" Rawls pasan a formar parte de la construcción positiva del contractualismo moral liberal. No obstante, la *Teoría de la justicia* de Rawls continúa siendo el punto de referencia polémico del contractualismo de Gauthier, porque a través de ese contraste se resalta no sólo que cabe partir de una posición inicial sin el artificio del velo de ignorancia (o con un velo más tenue), sino que necesariamente ha de ser así; pues ha quedado demostrado que las condiciones de los individuos rawlsianos implican la implausibilidad de los principios de la justicia que pretenden justificar.

Esperamos que este pormenorizado análisis del referente rawlsiano sirva para comprender mejor la insistencia de Gauthier sobre que las partes en la posición inicial son individuos *reales*, conscientes de que poseen capacidades, talentos, deseos e intereses propios, etc. Esta caracterización de las partes, unida a su concepción como individuos mutuamente desinteresados, instrumental o económicamente racionales y, en definitiva, conscientes de que tienen una identidad, configura el concepto global de individuo del que parte el contractualista liberal.

h) Individualismo y justificación.-

La caracterización expuesta de las partes (y sólo ella) permite que el contrato original no quede reducido a la decisión de un individuo abstracto en condiciones de extraordinaria incertidumbre, sino que sea un verdadero pacto, negociado entre agentes distintos que ponen su racionalidad instrumental al

[126] Gauthier, D., *MA*, p. 257; subrayado mío.

servicio de sus intereses. Es un proyecto ambicioso, tal vez cuestionable. Pero se trata de una potente re-definición del contractualismo hobbesiano (con la esperanza de que tenga éxito como filosofía moral) firmemente apoyada, por un lado, en el individualismo metodológico que hemos diseccionado y, por otro, en el análisis de la interacción racional tomado de la Teoría de la Decisión y Teoría de Juegos. Ambos dotan a la teoría de su plausibilidad inicial como expediente de justificación de las obligaciones morales.

Precisamente en relación con el objetivo último de la teoría, esto es, la justificación racional de las obligaciones morales, el individualismo contractualista reaparece con un nuevo sentido y una nueva cara, que queremos comentar antes de pasar al punto siguiente[127].

El contractualista moderno considera que únicamente el agente individual es sujeto de obligaciones morales. Sólo ante él han de ser justificadas dichas obligaciones. Él es, por así decir, el juez supremo en lo que concierne a la racionalidad (y, por tanto, a la obligatoriedad) de los deberes morales. Pero, ¿Quién es ese "agente individual" que hemos mentado como si fuera un universal? El término "agente individual" se refiere sencillamente a cada uno de nosotros, a las personas racionales que forman las sociedades y comunidades humanas. Desde luego, las personas somos distintas; esta convicción está en el centro de la ideología liberal. Es difícil encontrar un concepto de persona lo bastante abstracto como para que convenga a todas las concretamente existentes, tal es nuestra variedad, formas de vida y subjetividad moral, etc. Esto plantea un problema al teórico contractualista, comprometido con el reto de justificar las obligaciones morales universales ante los individuos concretos, pues ¿Qué tipo de razones pueden esperar el asentimiento de *cualquier* persona?

---

[127] Para aclarar el enfoque que adoptamos en este epígrafe, tal vez sea pertinente recordar la distinción entre los aspectos sustantivos y motivacionales de una teoría moral (Cfr. Jody S. Kraus y Jules L. Coleman, "Morality and the Theory of Rational Choice", *Ethics*, 97, Julio 1987, pp. 715-749; p. 715). Los comentarios anteriores pueden entenderse referidos a diversos aspectos sustantivos de la teoría, mientras el presente se centra en el aspecto motivacional, esto es, en el componente que "explica por qué una persona racional cumpliría los principios especificados por la teoría sustantiva" (Jody S. Kraus y Jules L. Coleman, *ibidem*).

La solución consiste en confiar en aquellas razones que apelan al substrato más elemental e innegable de las personas: a la racionalidad meramente instrumental que exhiben cuando eligen y deciden en vista de sus necesidades básicas. El contractualista se dirige, así, a individuos racionales despojados de los caracteres que les convierten en verdaderas personas. Estos individuos se pueden identificar a grandes rasgos con el constructo metodológico que hemos expuesto en las páginas anteriores.

Esta reducción puede parecer criticable. Si la moralidad ha de ser justificada ante las personas tal como son, ¿por qué se obvian de nuevo algunos de los constituyentes esenciales de la personalidad y se compone un discurso dirigido sólo a "hombres económicos"? La razón primordial es la siguiente: dado que somos diferentes, especialmente en lo que concierne en nuestros sentimientos hacia los demás, el único modo de construir un argumento válido para todos consistirá en ponerse en la situación hipotética más desfavorable; esto es, imaginar que nos hallamos en un mundo de hombres económicos e intentar construir aquél argumento justificatorio que tales individuos pudiesen aceptar. La postura del contractualista liberal es que la moral justificada ante el agente económico lo estará también, *a fortiori*, ante nosotros, individuos reales, miembros de sociedades y comunidades, que poseemos otros niveles de racionalidad y afectividad.

Por otro lado, esta convicción de que la moral justificada ante un individuo como el metodológicamente descrito lo estará también ante cualquier otro sujeto, permite que el contractualista liberal se mantenga fiel a la intención de no pre-juzgar ni definir a los individuos concretos. No es el objetivo de una moral por acuerdo ofrecer (ni aproximarse a) una definición "verdadera" de la subjetividad moral. Al justificar la moral ante el agente racional abstracto, pretendemos ofrecer un argumento válido para cualquier miembro de una sociedad libre y abierta, pero sin abrazar *a priori* tesis alguna sobre la naturaleza de la sociedad o de los individuos que la componen. Las fuentes de la subjetividad (incluso de la individualidad) no se discuten por el momento.

Podemos concluir entonces que el individualismo contractualista sigue siendo metodológico en su versión de "individualismo justificatorio". Pues se

admite que el individuo racional ante quien se despliega el argumento es una abstracción y una reducción. Pero una abstracción y reducción que permite mantener la neutralidad respecto al problema del carácter y origen de la subjetividad, sin renunciar a la normatividad de la teoría.

Estos límites metodológicos suponen un alto nivel de exigencia para el contractualista moral liberal, que acepta el reto de justificar racionalmente las obligaciones morales partiendo de un conjunto amplio de pre-juicios en contra de la moralidad. El cumplimiento de esta exigencia auto-impuesta debe garantizar la no-introducción de premisas espurias (ajenas a los conceptos de individuo, racionalidad, maximización, etc.). La pública confesión preliminar de los estrechos márgenes teóricos que el contractualista liberal se auto-concede nos convierte a todos en fiscalizadores de su rigurosa observancia.

El postulado del individualismo puede considerarse el primer presupuesto contra la moralidad. Al menos establece el marco inicial para la teoría, caracteriza a sus actores y, como acabamos de ver, anticipa el criterio que servirá para fijar los límites de su estricta racionalidad. Nuestro objetivo a continuación es completar el punto de partida del contractualismo liberal, mediante la precisión del modelo de racionalidad que se adopta como explicativo-descriptivo de la actuación de las partes en la posición inicial y, correlativamente, como instancia crítica de todo el despliegue argumental.

2. La racionalidad de las partes

a) Advertencias iniciales.-

Como hemos visto en el punto anterior, el contractualismo postula un individuo previo a la sociedad y ya dotado de ciertos atributos. Entre éstos, destaca la racionalidad. Suponemos que los individuos son agentes racionales, es decir, que existe un patrón de conducta compartido que permite prever cuáles serán sus acciones, dados sus intereses, preferencias o deseos y dado también un marco donde su elección no esté sometida a más restricción que la derivada de la escasez de los bienes capaces de satisfacer sus preferencias y la presencia de otros agentes que eventualmente compiten por los mismos bienes. Como punto de partida, este patrón se identifica con la racionalidad maximizadora tal como la entienden los economistas clásicos, fijada en la Teoría Bayesiana de la Decisión. El objeto de este epígrafe sobre la racionalidad de las partes es la explicación y discusión de algunas características de la racionalidad tal como es concebida por el neocontractualismo liberal; esta concepción se aparta en algunos puntos de la seguida por economistas y otros científicos sociales —de hecho, el contractualismo de Gauthier ofrece una revisión completa de la misma. La divergencia entre la teoría estándar y aquella adoptada y desarrollada por el contractualismo liberal justifica este epígrafe, pero a la vez limitará su alcance, pues en todo lo que el concepto de racionalidad se ajuste al representado por la Teoría de la Decisión, no será necesario que nos extendamos. No pretendemos, por tanto, explicar aquí toda la Teoría de la Decisión Racional; nos centraremos únicamente en los puntos más relevantes desde el punto de vista de la teoría moral contractualista. Nuestra exposición se desarrollará, en general, dentro del marco conceptual determi-

nado por las teorías de la Utilidad (que incluye la Teoría Bayesiana de la Decisión[128]) y de Juegos[129]. No obstante, creemos conveniente formular ahora algunas advertencias sobre nuestro enfoque particular, así como sobre el sentido de nuestra discusión sobre la racionalidad. Estas advertencias han sido sugeridas, al menos en parte, por los debates sostenidas en algunos seminarios con alumnos de quinto curso en la Facultad de Filosofía. No tienen otra misión que la de intentar adelantarse a ciertos malentendidos frecuentes, por lo que creemos que pueden ser obviadas por el lector familiarizado con la Teoría de la Decisión y el uso que de la misma se ha hecho en Filosofía Política y Moral.

El primer comentario se refiere a la relación entre la llamada "racionalidad económica" y la concepción de la racionalidad del contractualismo liberal. El modelo económico de racionalidad como maximización es, en efecto, adoptado por la Teoría de la Decisión y aplicado a muchas otras áreas de la ciencia social y política. La Filosofía Política y Moral también lo emplea, pero con ciertas modificaciones.

Gauthier reconoce que la visión de la racionalidad como maximización (propia de la economía) es demasiado vaga[130]; además, es ocioso decir que la especificación que de ella hace el economista está al servicio de la teoría

---

[128] Seguimos en este punto a John C. Harsanyi, y su división de las "disciplinas normativas sobre la conducta racional" expuesta en su artículo "Morality and the Theory of Rational Behaviour", en A. Sen y B. Williams, *Utilitarianism and Beyond*, Nueva York, Cambridge U. P., 1982, pp. 39-62. Una explicación sencilla y precisa del marco conceptual a que nos referimos puede verse en Gutiérrez López, G., "Racionalidad consecuencialista y restricciones deontológicas" (Muguerza, J. *et. al.* (eds.), *El fundamento de los derechos humanos*, Madrid, Debate, 1989, pp. 195-202), pp. 196-197.

[129] Como formulación académica canónica de estas teorías, tomaremos el libro de Michael D. Resnik, *Choices*, Minneapolis, University of Minnesota Press, 1987. La versión matemática axiomatizada y formal se encuentra en D. J. White, *Teoría de la decisión*, Madrid, Alianza, 1990, 3ª ed. (trad. de José Luis García Molina). Por último, sobre la Teoría de Juegos seguimos también a Morton D. Davis, *Introducción a la teoría de juegos*, Madrid, Alianza, 1986 (4ª reimpr.) (trad. José Carlos Gómez Borrero).

[130] Así lo escribe en "Economic Rationality and Moral Constraints" (*Midwest Studies in Philosophy*, III, 1978, pp. 75-96): "Kenneth Arrow habla de la 'tradicional identificación de la racionalidad con maximización de alguna clase', una formulación lo bastante vaga como para abarcar casi cualquier concepto".

económica: al concretar *quién* maximiza y *qué* maximiza, el economista introduce, por un lado, los personajes de su teoría, tales como la familia, la empresa, el Estado, etc.; y, por otro, el abstracto concepto de utilidad, como aquello que se trata de maximizar. Un uso filosófico de la racionalidad económica acepta —aunque reformulándolo— el concepto de utilidad (que es casi tautológico[131]), así como las condiciones formales y materiales para la definición de las funciones de utilidad[132]. Sin embargo, los individuos económicos (familias y empresas) han de ser sustituidos por otro tipo de personajes: sean las "personas", los "decisores individuales", o las "partes" del contrato; todos ellos comparten ciertos caracteres del *homo oeconomicus*, pero no se identifican con él. Por ejemplo, mientras el economista concibe al agente racional como un maximizador puro e irreflexivo o, con la frase frecuente,

---

[131] Cfr. D. Gauthier, "Economic Rationality and Moral Constraints", *cit.*, p. 76. En un artículo posterior a *MA*, Gauthier argumenta en contra de la relación necesaria que los economistas establecen entre decisiones tomadas por el agente, preferencia y utilidad (necesidad que conduce a una definición circular, pues la utilidad, que debe servir como criterio de elección, se define por relación a lo que se ha elegido). Nos referimos a "Economic Man and the Rational Reasoner", (en J.H. Nichols Jr. y C. Wright (eds.), *From Political Economy to Economics - And Back?*, San Francisco, Institute for Contemporary Studies Press, 1990, pp. 105-132), p. 125. El mismo argumento en contra de la visión económica de la decisión, se encontraba ya (algo menos explícito) en *MA*, p. 30: "Aceptamos el esquema explicativo general: la elección maximiza la satisfacción de preferencias dadas las creencias. Rechazamos la trivialización de este esquema que resulta de negar un acceso evidencial independiente a cada uno de sus términos —a la elección, la preferencia y las creencias. Cabe una elección desviada causada por una creencia errónea; cabe una elección irracional causada por una formación irracional de las creencias". Con esta tesis, Gauthier se aleja de la concepción económica, para la cual la coherencia en la ordenación de las preferencias es requisito suficiente de racionalidad. Como recogían Buchanan y Tullock (*The Calculus of Consent*, Ann Arbor, U. of Michigan P., 1962, p. 33): "...el economista moderno asume como hipótesis de trabajo que el individuo medio es capaz de clasificar u ordenar todas las combinaciones alternativas de bienes y servicios que pueden planteársele, y que ese orden es transitivo. Se dice que el comportamiento del individuo es "racional" cuando elige "más" en vez de "menos" y es coherente en sus eleciones". Para el economista, sólo una acción que no pueda considerarse coherentemente maximizadora (porque contraríe las condiciones formales para la definición de una función de utilidad) se reputará irracional. Desde el punto de vista de la Filosofía Política y Moral, por el contrario, será posible, en principio, hablar de irracionalidad en otros casos (errores en las creencias, etc.), aunque la acción responda formalmente a los requisitos de la racionalidad económica.

[132] Para una exposición sucinta de estas condiciones, ver Gauthier, "Economic Rationality and Moral Constraints", cit., pp. 76 y 77; también Gutiérrez, G., art. cit., p. 196. Para una explicación detallada, remitimos a Resnik, Michael D., *Choices*, Minneapolis, University of Minnesota Press, 1993 (3ª ed.), especialmente cap. 2.

como una "máquina de maximizar"[133], el filósofo moral —sea utilitarista o contractualista— debe contar, primero, con agentes cuya racionalidad incluya cierta capacidad auto-crítica y, segundo, con la posibilidad de tener un acceso directo y veraz a las preferencias y valores del sujeto —recordemos que el economista deduce las preferencias y valores del agente únicamente a partir de sus acciones coherentemente maximizadores[134].

Por todo ello, la concepción de la racionalidad de la que hablaremos, tiene que ver, pero no se identifica, con la llamada racionalidad económica. Su punto de referencia es más bien la Teoría de la Decisión, mucho más abarcante que la teoría económica, "producto de los esfuerzos conjuntos de economistas, matemáticos, filósofos, científicos sociales y estadísticos, encaminados a explicar cómo los individuos y los grupos toman o deberían tomar decisiones"[135].

En segundo lugar, calificaremos la racionalidad como instrumental. Harsanyi ha escrito que el concepto mismo de conducta racional nace del hecho empírico de que una gran parte del comportamiento humano consiste en acciones dirigidas a alcanzar fines[136]. Básicamente, la conducta racional es

---

[133] Esta concepción económica del agente racional ha sido criticada incluso por los mismos economistas (Cfr. D. Gauthier, "Economic Man and the Rational Reasoner", cit., p. 105). El modelo filosófico de agente racional no se identificó nunca con el *homo oeconomicus*, pero si alguna vez se acercó a él, el último Gauthier es especialmente crítico con dicho acercamiento, así ha escrito: "En tanto en cuanto traduce la realización global de los intereses de un individuo, que establece un problema de maximización-global, a un esfuerzo maximizador al nivel de cada elección particular, distorsiona nuestra comprensión del comportamiento racional" (Gauthier, D., "Value, Reasons and the Sense of Justice", en Frey, R.G. (ed.), *Value, Welfare and Morality*, Nueva york, Cambridge U.P., 1993, pp. 180-208; p. 204).

[134] A este respecto, puede verse la exposición "clásica" del modelo de racionalidad empleado por la ética y la teoría política (frente al empleado por los economistas) de Richard Brandt, "The Concept of Rationality in Ethical and Political Theory", en Pennock, J.R. y Chapman, J.W., *Human Nature in Politics*, Nueva York, New York University Press, 1977, pp. 265-279.

[135] M. D. Resnik, *op. cit.*, p. 3. Gauthier es explícito en este aspecto cuando escribe: "Doy por sentado que la racionalidad posee un contenido sustantivo no arbitrario. Este contenido es captado, o intenta serlo, por la Teoría de la Decisión Racional." ("Economic Man and the Rational Reasoner", en Nichols, J.H. y Wright, C. (eds.) *From Political Economy to economics —And Back?*, San Francisco, Institute for Contemporary Studies Press, 1990, pp. 105-132; p. 108).

[136] Cfr. Harsanyi, J.C., art. cit., p. 42.

aquella que resulta coherente con la búsqueda de algunos objetivos bien definidos de acuerdo con un conjunto de preferencias y prioridades asimismo bien definidas. La conducta racional es un medio para la satisfacción de los deseos, intereses o preferencias individuales y colectivas.

Abundaremos en esta idea más abajo, pero en relación a ella creemos importante recordar lo siguiente: que el hecho de asumir el concepto instrumental de racionalidad usado por la Teoría Bayesiana de la Decisión no implica, al menos en principio, *negar* que exista o que pueda existir otro tipo de racionalidad "superior". Cuando Gauthier escribe, sobre el decisor ideal, que "su racionalidad se expresa en el esfuerzo por maximizar la satisfacción de sus preferencias, dadas sus capacidades y otros rasgos de su carácter en las circunstancias en las que se encuentre, cualesquiera que sean", y que "*No hay ningún otro nivel de racionalidad implicado*"[137], se apresura a explicar que el sentido de sus palabras es que, aunque algún otro "nivel" de racionalidad pueda ser útil para otros propósitos, sería irrelevante para una teoría de la moral basada en la decisión racional. En mi opinión, estas palabras equivalen a decir que, si bien es imposible negar que haya niveles "sustantivos" (y no meramente instrumentales) de racionalidad, el contractualista moral *no los necesita*, pues considera que su método conseguirá fundamentar una moral sin recurrir a tales premisas. Que haya una realidad que corresponda a alguno de los conceptos de racionalidad que se presentan como alternativos a la racionalidad instrumental no puede negarse apodícticamente. Sencillamente, el contractualista moral mantiene que no necesita de tales hipótesis, y que la carga de la prueba pesa sobre aquél que las defienda.

Mientras la presencia de "otros" niveles de racionalidad, o la existencia de una "razón común", son hipótesis enormemente problemáticas, la racionalidad instrumental inherente a nuestra esencia de seres perseguidores de objetivos es innegable. Ante esto, el contractualista opta por intentar construir su argumento adoptando una premisa evidente, en vez de otras oscuras y cuestionables, que demandarían una compleja justificación. Pero tal opción (que podemos denominar "la opción por la debilidad de las premisas") no implica

[137] *MA*, p. 256; subrayado mío.

la negación apriorística de otras concepciones de la racionalidad. De hecho, la conclusión del contractualismo moral es que *hay* un razonamiento genuinamente moral que surge de esta racionalidad instrumental simplemente maximizadora, pero que difiere de ella y la supera.

Por último, precisaremos algo sobre el sentido de la racionalidad instrumental como maximización. Puesto que se trata de conceptos abstractos, conviene aclarar someramente qué significa "maximizar" y "qué es aquello que se maximiza". Nos centraremos, por tanto, en los conceptos de utilidad y preferencia, y su contenido posible[138].

Para la teoría económica, los agentes racionales maximizan (o deben maximizar) su utilidad individual. La utilidad es una medida de la satisfacción de las preferencias, luego maximizar la utilidad no significa otra cosa que satisfacer en el mayor grado posible las preferencias individuales. Así, escribe Gauthier que "la persona racional trata de maximizar la satisfacción de sus

[138] Un análisis profundo de la concepción gautheriana de la preferencia puede verse en Kurt Baier, "Rationality, Value and Preference" (en E. F. Paul *et al.* (eds.), *The New Social Contract,* Oxford, Blackwell, 1988, pp. 17-45), especialmente secciones III-VIII. Estamos en general de acuerdo con el análisis de Baier sobre los dos aspectos clave de la concepción de las preferencias expuesta en *MA*, a saber, la distinción (que los economistas obvian) entre preferencias reveladas en las elecciones, y actitudes (o preferencias expresadas previamente o al margen de la elección), y la necesidad de que las preferencias sean reflexivas y completamente meditadas. Baier es extraordinariamente crítico, sin embargo, con la visión de Gauthier, acusándole, en relación con el primer aspecto, de no haber logrado superar la concepción económica (anclada en las preferencias "reveladas en la conducta") y, en relación con el segundo, de necesitar un modelo de racionalidad práctica previo para explicar sobre qué base se "reflexionan o meditan" las preferencias. Tal "racionalidad práctica" interviniente en la formación de preferencias consideradas vendría a quebrar por completo el proyecto contractualista de Gauthier, pues habría de contener criterios normativos previos al acuerdo. En nuestra opinión, Baier exagera el alcance de la concepción de las preferencias y realiza una crítica pretendidamente radical allí donde la única debilidad de Gauthier reside en que evita un prolijo debate y justificación de sus tesis. De este modo, la respuesta (creemos que satisfactoria) a las críticas de Baier puede encontrarse en los artículos recientes ("Economic Man and the Rational Reasoner", "Value, Reasons and the Sense of Justice" y "Assure and Threaten", de 1990, 1993 y 1994 respectivamente), en los que Gauthier profundiza en un enfoque de las preferencias y el valor cada vez más apartado del modelo económico que predomina en *MA* y muestra —empleando ideas como la "coherencia con los estados intencionales previos", la "unificación de los estados intencionales en una experiencia individual", etc.— cómo se forman las preferencias meditadas a partir del concepto de "finalidad" (entendida como proyecto personal global) de la acción.

preferencias meditadas, cualquiera que sea su contenido"[139]. Estas últimas palabras, "cualquiera que sea su contenido", provocan incomprensiones y dudas. Según una idea pre-teórica de racionalidad, no consideraríamos racional la conducta de alguien cuyas preferencias tuviesen un contenido excesivamente "inusual" (por ejemplo, la preferencia que un suicida tiene por su propia muerte); sin embargo, de acuerdo con el concepto de racionalidad que emplearemos, el sentido de las palabras de Gauthier es el literal. La Teoría de la Decisión parte de la idea de que toda decisión individual está basada en razones subjetivas, cuyo contenido viene dado por preferencias asímismo subjetivas y, por lo tanto, no criticables desde el paradigma de la racionalidad instrumental. El hecho de que ciertos estados de cosas relativos al agente sean muy mayoritariamente preferidos (o no-preferidos) se trata como una contingencia sin relevancia normativa: una preferencia (p. ej., la de permanecer vivo) no es más racional por el hecho de ser mayoritaria (de hecho, el adjetivo "racional" estaría impropiamente aplicado a cualquier preferencia). La correcta comprensión de la racionalidad como instrumental supone que el contenido de las preferencias es estrictamente irrelevante para el análisis (siempre que se trate de preferencias meditadas y fundadas en creencias e informaciones completas y veraces)[140].

Ello tiene una consecuencia un tanto inesperada. Como hemos visto en el punto anterior, se postula que los individuos son "mutuamente desinteresados". Pero si el contenido de las preferencias no influye en la racionalidad de la conducta, cabría albergar racionalmente preferencias altruistas, lo que parece contradecir el postulado del desinterés mutuo. Mas es esta una contradicción que se revela falsa si tenemos en cuenta que el desinterés mutuo se refiere únicamente a cómo se conciben las partes del contrato, no a cómo es la racionalidad instrumental en general. No obstante, esta falsa contradicción

---

[139] Gauthier, D., "Reply to Wolfram" (en *Philosophical Books*, vol. 23, nº 3, julio 1987, pp. 134-139), p. 135.

[140] En este punto, el contractualismo moral liberal sigue fielmente la conocida tesis humeana sobre la incapacidad de la razón para decidir sobre fines últimos. Cfr. David Hume, *A Treatise of Human Nature*, libro II, parte III secc. III (pp. 114 y ss. de la edición de L.A. Selby-Bigge y P. H. Nidditch, Oxford, Clarendon, 1978).

puede dar lugar a malentendidos, por lo que Gauthier creyó conveniente afirmar explícitamente:

> "*Para los propósitos de mi argumento*, no asumo preferencias orientadas a otros; trato de mostrar que la moralidad es compatible con el desinterés mutuo (...). Pero esta concepción de las personas *no forma parte de la racionalidad*".[141]

El único límite en el contenido de las preferencias (y, por tanto, en la concepción de la racionalidad) es que las preferencias respondan siempre al interés del agente. El auto-interés (como lo denominaremos en adelante) implica que la racionalidad se orienta a la satisfacción de las preferencias *del* agente que prefiere, es decir, a la maximización de *su* utilidad. En este sentido, habría que matizar el uso del término "altruismo". Si se entiende como lo hemos hecho arriba, con el significado común de "preferencia *de* un agente que consiste en querer cierto estado de cosas (positivamente valorado) *para otro u otros* agentes", el altruismo es concebible y admisible como uno de los contenidos posibles de las preferencias o intereses individuales (aunque parezca contradecir el postulado del "desinterés mutuo"). El contenido "altruista" o "benéfico" de una preferencia sería tan irrelevante, desde el punto de vista de la teoría de la racionalidad instrumental, como el contenido de las preferencias suicidas de un suicida.

Sin embargo, se excluye como contenido posible de las preferencias racionales un "altruismo" que, como el defendido por Nagel, no pudiera ser "traducido" a términos de maximización del interés del agente que elige y actúa[142]. Un altruismo no reconducible a auto-interés, entendido como

---

[141] Gauthier, D., "Reply to Wolfram", cit., p. 135; subrayado mío.

[142] Es decir, un altruismo que no pudiera explicarse por referencia al auto-interés del agente, sino que definiese aquella estructura de preferencias de un agente que se identifica o contiene *la preferencia de* otro agente (de modo que las preferencias del primer agente se niegan a sí mismas); este concepto de "altruismo" no puede admitirse como un tipo de preferencias, pues su contenido contradice el sentido mismo de una preferencia subjetiva. El altruismo a que nos referíamos en el párrafo anterior, que podemos identificar como "benevolencia", no presenta problemas: cabe que un agente prefiera satisfacer las preferencias de un tercero (independientemente de que,

abnegación o renuncia a las propias preferencias, quizá se dé (aunque es muy cuestionable[143]), pero supondría, no sólo un grado de benevolencia con el que es ilegítimo contar en este nivel del argumento, sino un tipo de racionalidad diferente, denominada "universalista" por Gauthier[144], que necesitaría ser probada, frente a la mayor evidencia de la concepción instrumental maximizadora[145].

---

probablemente, la idea misma de benevolencia tenga su raíz en experiencias subjetivas). Tal tipo de preferencia no implica negar que el agente sea auto-interesado, ya que es su propia preferencia (benevolente) la que satisface al procurar satisfacción al tercero.

Para aclarar la diferencia entre altruismo como benevolencia (admitido como contenido de las preferencias individuales) y altruismo como abnegación (excluido), tal vez no sea ocioso recordar el concepto de altruismo que emplea Nagel en *The Possibility of Altruism* (Oxford, Clarendon, 1970). Nagel considera que los intereses de otras personas pueden ser motivos para la acción directamente, "sin necesidad de motivos ulteriores" (p. 79). Esto supone que la razón práctica de los agentes es algo así como un mecanismo automático, que funciona por respuesta a intereses, sin importar de quién sean éstos. Según Nagel, esta concepción de la razón práctica permite el altruismo. Según la teoría de la decisión, es sencillamente incorrecta. Toda acción exige un motivo subjetivo, o, al menos, radicado en el interés del agente que actúa, por más generoso y benéfico que tal interés sea. Desde el punto de vista de la Teoría de la Decisión Racional, un altruismo como el que defiende Nagel es incoherente con la percepción de la racionalidad como maximización de las preferencias individuales y haría imposible la definición precisa de funciones de utilidad personales.

[143] Tal abnegación supondría, para una concepción económica del individuo, la negación de la propia personalidad, ya que nuestros deseos y preferencias (aunque se refieran a otros) nos identifican como individuos distintos y únicos. Desde este punto de vista, el altruista abnegado, que pone en lugar de sus preferencias las de otros, deja de ser un individuo distinguible; pasa a ser un medio para fines ajenos (como exige, por cierto, la "concepción universalista" de la razón defendida por el utilitarismo).

Pese a estas dudas, no hay que confundir el altruismo del que hablamos, cuya posibilidad o imposibilidad es un problema filosófico, con la cuestión técnica de la independencia de las funciones de utilidad individual, considerada requisito necesario para poder establecer el modelo de un mercado perfectamente competitivo y, por tanto, normativamente defendida por la economía clásica (cfr. Gauthier, *MA*, pp. 86-87). Sobre esta cuestión bastarán las ideas intuitivas que expusimos en el punto anterior (cfr. arriba pp. 85 y 86, en especial la tabla en la nota 79).

[144] Conectada, por ejemplo, con la idea de "razones objetivas" de Nagel.

[145] Cfr. Gauthier *MA* p. 7. El contraste entre ambas concepciones de la racionalidad queda claro en el siguiente párrafo: "Para evitar posibles malentendidos, téngase en cuenta que ninguna concepción de la racionalidad requiere que las razones prácticas sean egoístas. Según la concepción maximizadora no es el interés en el yo (*self*), tomado como objeto, el que proporciona una base para la decisión y acción racionales, sino el interés del yo (*self*), considerado como sujeto. Según la concepción universalista, no son los intereses en cualquiera, tomando a todas las personas como objetos, sino el interés de cualquiera, tomado como sujeto, el que proporciona una base para la decisión y la acción racionales. Si yo tengo un interés directo en tu bienestar, entonces, según cualquiera de estas dos concepciones, tendré una razón para promoverlo. Pero tu interés en tu

De todo lo dicho se desprende que la maximización del auto-interés no debe juzgarse peyorativamente. Simplemente describe el hecho de que cada agente intenta maximizar *sus* preferencias, cualesquiera que estas sean. Ser auto-interesado (*self-interested*) no implica ser egoísta (*selfish*).

No obstante, la distinción entre el egoísta y el agente maximizador auto-interesado no aparece, en general, clara. Con frecuencia ambos términos se emplean indistintamente. Por ello, tal vez convenga resaltar la diferencia, al menos a efectos de este trabajo: hablaremos en general de agentes auto-interesados para referirnos al agente instrumentalmente racional que persigue la satisfacción de sus preferencias; mientras que reservaremos el término "egoísta" para aquellos agentes cuyas preferencias les tienen a ellos mismos (su bienestar, placer, etc.) como objeto. El término "egoísmo" queda restringido, por tanto, a la descripción de una actitud psicológica o ética contingente, mientras que el "auto-interés" es un carácter esencial de la racionalidad, según la concepción maximizadora de la misma.

Como señalábamos, esta distinción no es siempre respetada. En particular, entre los teóricos de juegos es común hablar de "estrategia egoísta" para hablar de la estrategia o respuesta racional maximizadora auto-interesada. Del mismo modo, se suele hablar de "estrategia altruista" para referirse a cualquier respuesta distinta de la recomendada por la maximización de utilidad. Ambos términos "egoísta" y "altruista" se emplean convencionalmente por simple comodidad. Es evidente que, según nuestra terminología, se refieren, respectivamente, a estrategias auto-interesadas o no auto-interesadas.

Estas precisiones iniciales no agotan los problemas, dudas y ambigüedades que acompañan al intento de defender, desde un punto de vista filosófico, una concepción de la racionalidad un tanto magra y, si bien familiar para los científicos sociales, ajena a las tradiciones filosóficas continentales. Esperamos que, aunque sea imperfecta y parcialmente, contribuyan a una comprensión mejor de lo que expondremos a continuación. En nuestra exposición

---

propio bienestar me proporcionará una razón sólo según la concepción universalista". (De las últimas frases concluimos que la concepción universalista equivale, más o menos, a la idea del altruismo defendida por Nagel en *The Possibility of Altruism*).

—siguiendo nuestra intención de no abundar innecesariamente en aquellas notas definitorias del concepto de racionalidad que ni presentan problemas ni especiales peculiaridades en su versión contractualista— nos centraremos en los conceptos de instrumentalidad, maximización, y en la articulación de la Teoría de la Decisión Racional con una teoría moral contractualista. Por último, haremos una breve referencia a las concepciones de la racionalidad (y su papel en relación con la teoría moral) que pueden considerarse alternativas a la que aquí proponemos, siguiendo principalmente a Gauthier.

b) Racionalidad instrumental, consecuencialismo y prudencia.-

Al inicio mismo del presente capítulo reproducíamos un texto de Popper que establece, con rara precisión, el método contractualista. Es éste un método de construcción racional —decía Popper—, basado en una suposición de *completa racionalidad*. Nos toca ahora aclarar qué concepto de racionalidad suponen los contractualistas que describe mejor la conducta de los individuos; qué significa que un individuo sea "completamente racional".

Para el padre de la tradición contractualista moderna, Thomas Hobbes —cuya teoría no difiere, metodológicamente, del contractualismo contemporáneo—, la razón (objeto del capítulo quinto del *Leviatán*) se reduce a cálculo. En los asuntos prácticos, el cálculo se identifica con la búsqueda y conocimiento de las consecuencias de las cosas y los efectos que ciertos actos pueden lograr mediante ellas, así como con la reducción de dichas consecuencias a reglas generales llamadas *teoremas* o *aforismos*. Como la aritmética, es el razonar un arte que hay que aprender, perfeccionar e intentar usar correctamente, de acuerdo a reglas precisas. Pero a diferencia de la aritmética, no proporciona certeza. Porque si bien todos los hombres poseen una razón (más o menos desarrollada y fiable), no están dotados por naturaleza de *una recta razón* capaz de decidir inapelablemente las disputas entre ellos. Así, rige para la racionalidad un principio de estricto individualismo: cada cual razona, pero

en sus cálculos siempre dominan las pasiones, en definitiva, el egoísmo.

En esta primera aproximación la razón aparece, pues, como una máquina de calcular puesta al servicio de los deseos o necesidades (egoístas) del agente[146].

La misma idea básica sobre la razón (excepto por lo que se refiere al egoísmo) reside en la filosofía de Hume. En el *Treatise*, es defendida con argumentos cada vez más convincentes y potentes, hasta concluir que

> "La razón es y sólo debe ser esclava de las pasiones, y no puede jamás pretender otro oficio que el de servirlas y obedecerlas."[147]

Como en el caso de Hobbes, pero con mayor sutileza, la razón es concebida por Hume como una máquina de calcular efectos, inferir consecuencias y comparar ideas. Las operaciones de la razón pueden "señalar" a las pasiones las consecuencias de ciertos objetos o estados de cosas[148], pero el impulso que gobierna la voluntad tiene siempre su origen en las pasiones. La razón calculadora está radicalmente separada de la acción:

> "El razonamiento abstracto o demostrativo no influye nunca en ninguna de nuestras acciones, salvo en la medida en que dirige nuestro juicio sobre las causas y los efectos."[149]

---

[146] Incluso cuando, al final del capítulo XIII, Hobbes escribe que la razón ha de sugerir los convenientes artículos para la paz (y superar así los males del estado de naturaleza), tiene la razón un papel instrumental, al servicio de "las pasiones que inclinan a los hombres a la paz, [que] son el miedo a la muerte; el deseo de las cosas necesarias para una vida confortable; y la esperanza de obtenerlas mediante su industria".

[147] Hume, D., *A Treatise of Human Nature*, Oxford, Clarendon, 1978 (ed. de L.A. Selby-Bigge), Libro II, Parte III, Sección III (p. 415).

[148] Y aún en esta labor, no puede ofrecer la razón certidumbre pues, de los tres grados de evidencia que puede alcanzar la razón (Cfr. Hume, *Treatise*, Libro I, Parte III, Sección XI, p. 124), sólo el más débil, el fundado en la probabilidad, se da en las materias prácticas.

[149] *Ibidem* (p. 414).

La razón es hasta tal punto inconmensurable con las pasiones y la acción, que es incapaz de condenarlas o justificarlas, salvo en cuanto implican actos de juicio. Así, una pasión o un acto pueden considerarse irracionales sólo si se encuentran fundados en creencias erróneas o suponen una incorrecta elección de medios para el fin que pretenden. En todos los demás casos, nada tiene la razón que decir sobre el contenido de las preferencias[150].

La visión humeana de la razón resulta, en conclusión, poderosamente simple —tanto, que no difiere de la "razón de los animales"[151]— pero, a cambio, totalmente inútil para la ética:

> "La moral excita las pasiones, y produce o prohíbe acciones. La razón por sí misma es completamente impotente en este particular. Las reglas de la moralidad no son, por lo tanto, conclusiones de nuestra razón."[152]

Y sin embargo, ¿qué otra cosa sino conclusiones de la razón podrían ser unas reglas que generalmente se oponen a nuestros deseos y establecen límites a nuestras inclinaciones? Como dice Gauthier al comienzo de *MA*,

> "Si las demandas de la moral han de tener algún efecto práctico, alguna influencia sobre nuestro comportamiento, no es porque susurren seductoras a nuestros deseos, sino porque convencen a

---

[150] Cfr. *Ibidem* (p. 416).

Este punto ha sido magníficamente explicado por Baier, cuya versión de la racionalidad instrumental humeana exponemos: "la razón es algo así como una calculadora mental que nos permite resolver cosas tales como las conexiones causales entre los sucesos y, por tanto, también entre nuestras acciones y sus consecuencias. Así, la razón nos permite deducir qué debemos hacer para lograr ciertos fines. Según este punto de vista, determinamos lo que es conforme a la razón y contrario a ella especificando cuáles son los medios apropiados para nuestros fines [...]. En lo que concierne a nuestros fines mismos, la razón no tiene nada que decir,porque nuestros fines no están determinados por la razón, sino por las pasiones. De ahí que la cuestión sobre si nuestros fines son conformes o contrarios a la razón simplemente no puede plantearse." (Baier, K., *The Moral Point of View*, Ithaca, Cornell U.P., 1958, p. 90).

[151] Cfr. *Treatise*, Libro I, Parte II, Sección XVI (pp. 176 y ss.).

[152] Hume, D., *Treatise*, Libro III, Parte I, Sección I, p. 456.

nuestro intelecto."[153]

Como quiera que Hobbes fue pionero en el intento de extraer consecuencias prácticas de la racionalidad instrumental auto-interesada, a él y su teoría volvieron los ojos quienes, seguros de que la visión humeana de la racionalidad era esencialmente correcta, no se resignaban a aceptar su impotencia práctica. Nace así una tradición de pensadores cuyo objetivo es demostrar que cabe derivar conclusiones normativas (éticas o políticas) a partir de un concepto instrumental de racionalidad; y cuyo método sigue, más o menos de cerca, al iniciado por Hobbes.

Este intento por mantener la influencia en la esfera práctica de una razón concebida instrumentalmente, por evitar la reducción de la ética a compasión y de la política a benevolencia, adopta dos formas básicas: el utilitarismo y el contractualismo. Ambas tradiciones tratan, por caminos diferentes, de mostrar la racionalidad (es decir, la utilidad) de las restricciones deontológicas.

Sin embargo, sólo el contractualismo es rigurosamente fiel a la concepción instrumental de la racionalidad. El utilitarismo confía, ciertamente, en el papel instrumental de la razón, que nos informa sobre los medios para alcanzar nuestros fines, pero en la medida en que dicho fin (la utilidad) se identifica con algún tipo de bien o medida objetiva cognoscible, la razón pierde su carácter exclusivamente instrumental. Este movimiento es evitado tanto por los contractualistas clásicos como por los contemporáneos, pues todos ellos permanecen dentro de los límites del modelo humeano de razón enteramente separada de sus fines, sin introducir *a priori* concepto objetivo alguno de "lo bueno" o "lo útil". Esta fidelidad secular significa en el presente la aceptación, en principio, de la Teoría de la Decisión Racional como descripción estándar de la racionalidad, y el rechazo de todo componente teleológico en la misma; en palabras de Gauthier:

[153] Gauthier, D., *MA*, p. 1.

> "*La Teoría de la Decisión Racional también trata la razón práctica como estrictamente instrumental*. Esto no está implícito en la identificación de racionalidad con maximización porque, como hemos señalado, se podría suponer que la cantidad a maximizar no fuese una medida, sino un modelo de preferencias, un valor objetivamente inherente a estados de cosas, cuya aprehensión dependería del ejercicio de la razón. Desde este punto de vista, la razón no sería simplemente instrumental, sino que tendría que ver también con los fines de la acción. Sin embargo, al identificar la racionalidad con la maximización de una medida de la preferencia, *la Teoría de la Decisión Racional rechaza toda conexión con los fines de la acción. Los fines se infieren de las preferencias individuales*; si las relaciones entre estas preferencias y el modo en que se mantienen satisfacen las condiciones de la decisión racional, entonces la teoría acepta cualesquiera fines que impliquen."[154]

La fidelidad al marco de la Teoría de la Decisión supone, por lo tanto, aceptar los límites de la racionalidad instrumental; situar a la razón en su modesto lugar. Pues, como ha repetido Gauthier más recientemente, "es un error suponer que la razón es el árbitro último de si los objetos de preferencia son apropiados o inapropiados [...]. No debemos asignar a la razón tareas inadecuadas para su categoría, tales como [...] elegir entre preferencias"[155].

Esta concepción "modesta" de la racionalidad ha sufrido, a lo largo de su historia, muchas críticas y más incomprensiones. No hay aquí espacio para intentar persuadir a quienes dudan de sus posibilidades filosóficas (o, más bien,

---

[154] Gauthier, D., *MA*, p. 25-26 (subrayado mío). Las principales condiciones de la decisión racional a que se refiere el texto son las condiciones estándar de coherencia en los órdenes de preferencia: tratarse de preferencias consideradas, formar un orden completo, transitivo y continuo.

[155] Gauthier, D., "Economic Man and the Rational Reasoner", en J.H. Nichols Jr. y C. Wright (eds.) *From Political Economy to economics—And Back?*, San Francisco, Institute for Contemporary Studies Press, 1990, pp. 105-133; pp. 109-110.

el trabajo entero es el argumento), pero sí queremos hacernos cargo de, al menos, una de las posibles críticas contra ella. Se trata de la formulada por Nagel, representante contemporáneo de una visión objetivista de la racionalidad, hace ya algún tiempo. Reproducimos por extenso una de sus versiones:

> "El punto crucial es que una razón práctica es una razón para hacer o querer algo, como una razón teórica es una razón para concluir o creer algo. Este es el carácter definidor del razonamiento práctico. Mantener, como Hume, que la única crítica posible a la acción es una crítica a las creencias asociadas a ella, es mantener que la razón práctica no existe. Si reconocemos la existencia de razones para la acción debemos mantener no sólo que ellas justifican que creamos ciertas proposiciones sobre la acción, sino que ellas justifican la acción misma. Debe ser posible comprometerse en una crítica racional de la acción y el deseo, y no sólo de las creencias asociadas a ellos."[156]

La crítica de Nagel está bien fundada en su análisis de la conexión entre motivación, razón y acción. Dirigida contra Hume, tiene éxito. Sin embargo, no lo tiene frente al contractualismo, pese a que éste coincida con Hume en la aceptación de una racionalidad ajena a los fines. Ello se debe a que el argumento contractualista intenta, en definitiva, precisamente aquello que Nagel reclama: comprometerse en una crítica racional de la acción y el deseo. Lo que ocurre es que las razones para la acción abarcan un espectro muy amplio de hechos, creencias y normas, algunas de las cuales requieren una compleja justificación ante el individuo. El contractualista es consciente de esto y, a los efectos de esa justificación, y como concepto básico, concibe la racionalidad separada de (aunque relacionada con) las preferencias subjetivas y completamente ajena a unos hipotéticos fines objetivos. Mas una vez justificadas individual e instrumentalmente, las normas, hechos y creencias que constituyen razones para la acción se convierten en impulsos motivadores que conforman

[156] Nagel, T., *The Possibility of Altruism*, Oxford, Clarendon, 1970, p. 64.

una densa racionalidad práctica reflexiva y crítica. Como conclusión de su esfuerzo, el contractualista espera superar el pobre estatuto en que Hume dejó a la razón (y satisfacer, en cierto modo, las demandas de un objetivista como Nagel); espera, además, hacerlo partiendo de una radical presunción en contra del objetivismo, utilizando como base, el magro e indiscutible hecho empírico de que somos unos seres con necesidades, carencias y deseos, cuyas acciones (la inmensa mayoría de ellas) están encaminadas a satisfacerlos. El uso instrumental de la razón es el único inapelablemente fundado en la experiencia. Si a partir de este uso, una parte de nuestro comportamiento (aquella que no tiene que ver con el logro de objetivos auto-interesados, sino con el cumplimiento de normas) no pudiera ser explicada, entonces habría que concluir, bien con Hume, negando toda posibilidad de fundar racionalmente las reglas morales, bien con Kant, considerando que la norma moral es un imperativo indisociablemente unido a la razón (postulando, por tanto, una razón preñada de moralidad). Pero si, como el contractualista defiende, es posible explicar *también* el comportamiento conforme a reglas desde la racionalidad instrumental, entonces, tanto el emotivismo, como una concepción "moral" de la razón, estarían injustificados.

La idea que subyace al contractualismo liberal es que, aunque la razón instrumental está vacía de contenido, impone un modo de interacción en cuya estructura es posible hallar el fundamento para una revisión racional del comportamiento meramente auto-interesado. Es decir, las acciones conforme a reglas no son un hecho inexplicable desde el punto de vista de la racionalidad instrumental, sino que, muy al contrario, vienen requeridas por ella misma. Mediante el argumento contractualista se nos mostrará cómo la estructura de la interacción entre agentes auto-interesados da lugar a un conjunto de restricciones mutuas a las que resulta individualmente racional someterse y que, en última instancia, conforman una estructura de fines y valores compartidos.

Queremos que estos comentarios, que adelantan conclusiones mientras posponen su defensa, sirvan para justificar la aceptación del concepto instrumental de racionalidad, al exponer cómo sería posible escapar a las críticas vertidas contra la visión humeana. Esa posibilidad será real, desde luego, sólo si la teoría demuestra lo que promete.

Por otro lado, el modelo de racionalidad maximizadora y utilitaria que presenta la Teoría de la Decisión, une a sus virtudes y posibilidades el inconveniente de oscurecer la simple idea de que la racionalidad instrumental no es algo esencialmente diferente de la prudencia, entendida, en sentido kantiano, como el tipo de deliberación mundana encaminado a decidir sobre los medios más adecuados para alcanzar una vida feliz[157]. La única precisión que tal vez habría que hacer, una vez más, es que para la Teoría de la Decisión, dicha "vida feliz", como fin prudencial, tiene un contenido subjetivo (los fines son establecidos por cada agente) y no criticable. Mas es ésta una precisión que no debería extrañar a la vista de las palabras de Aristóteles, quien ya reconoció que, en cuanto deliberación práctica, la razón no se interesa por los fines de la acción:

> "...el hombre es el principio de las acciones, y la deliberación tiene por objeto lo que él mismo puede hacer, y las acciones se hacen en vista de otras cosas. Pues no puede ser objeto de deliberación el fin, sino los medios conducentes a los fines [...] El objeto de la deliberación y de la elección son el mismo, salvo que el de la elección está ya determinado, pues se elige lo que se ha decidido como resultado de la deliberación"[158]

Ciertamente, se puede argumentar, en relación con el punto de vista aristotélico, que aunque la razón instrumental deliberadora está subordinada a los fines particulares, éstos son establecidos por el *telos* de cada miembro de la comunidad ("lo que es bueno hacer y ser para alguien como él") que, una vez identificado por la razón teorética, debe guiar la edu-cación de las pasiones. Según esta visión, la presencia en la deliberación de los deseos y metas del agente, no responde a un origen subjetivo no-criticable, basado en sus preferencias libérrimas, sino que tiene un fundamento teleológico objetivo que

---

[157] Cfr. Kant, I., *Fundamentación de la metafísica de las costumbres*, Madrid, Espasa-Calpe, 1990 (9ª ed., trad. de Luis Martínez de Velasco), cap. II, p. 86 y ss.

[158] Aristóteles, *Ética a Nicómaco*, 1112*b*-1113*a*.

da pie a juicios críticos sobre las acciones e incluso sobre los deseos y metas del agente concreto[159].

Sin entrar a discutir la precisión de las interpretaciones aristotélicas del tipo de la que hemos aludido, sería indudablemente anacrónico atribuir a la concepción aristotélica de la razón práctica los conceptos liberales que caracterizan a las explicaciones contemporáneas. Su referencia al papel de la deliberación y su relación con los fines nos sirve simplemente como certificación de que la idea de instrumentalidad está asociada a nuestra comprensión más inmediata de la racionalidad práctica.

Una prueba aún más significativa de esto mismo es que un objetivista actual, como Nagel, admita que el origen de las razones prácticas está en los deseos y carencias del agente, que imponen los fines; éstos, a su vez proporcionan razones para ciertas acciones[160]. Tal como interpreto la idea de Nagel, las razones para la acción, aunque devengan lógicamente objetivas, tienen un origen radicado en el sujeto que desea y actúa, y en las relaciones mundanas entre medios y fines auto-interesados.

Como vemos, el componente instrumental de la racionalidad, legítimamente identificable con la deliberación práctica prudencial, conecta con tradiciones filosóficas no necesariamente asociadas al empirismo humeano.

El rasgo distintivo de la racionalidad asociada a la Teoría de la Decisión es su peculiar interpretación de la acción, basado en el concepto de "preferencia individual". Convertir las preferencias y decisiones individuales en el punto de apoyo de la concepción de la racionalidad singulariza esta concreta visión de la razón instrumental. La elección de conceptos menos inmediatos, como los de "interés", "meta", "objetivo", etc. y el intento de explicarlos sin referencia a las preferencias, conduce a Nagel, por ejemplo, a un objetivismo demasiado simplista. Por otro lado, la creencia de que la razón está al servicio de la mera pasión (llamemosla "deseo", "impulso" o con cualquier otro nombre) condujo

---

[159] Cfr. MacIntyre, A., *Tras la virtud*, Barcelona, Crítica, 1987 (trad. Amelia Valcárcel), pp. 203-204.

[160] Cfr. Nagel, T., *The Possibility of Altruism*, cit., p. 33 y ss.

a Hume a no distinguir nuestra razón de la de los animales.

El concepto de preferencia media entre los impulsos o necesidades y los objetivos del agente. Para concretar este concepto partimos, ciertamente, de las necesidades inmediatas (instintivas o biológicas, podríamos decir) pero reconocemos en el ser humano capacidades que permiten un modelo de acción no equiparable al de la mayoría de los animales, que responden automáticamente a sus instintos[161]. Nuestra capacidad (compartida, dicho sea de paso, con algunos otros animales) de concebir creencias, sentir emociones, alimentar anhelos, esto es, el hecho de que nuestros actos mentales posean un contenido intencional, posibilita que nuestra acción sea *elegida*. Mas la mera intencionalidad no es aún suficiente para aplicar las categorías de análisis de la Teoría de la Decisión; es necesario, además, la capacidad de representación semántica de los estados intencionales mismos[162], y esta capacidad sí es, que sepamos, exclusivamente humana[163]. Gracias a esta capacidad podemos figurarnos y comparar estados de cosas y, al hacerlo, se ocasionan muchas veces incoherencias y contradicciones. En esos casos, la razón se impone. Unas veces, porque los estados de cosas que imaginamos son incoherentes con nuestras propias creencias. Otras, porque representan situaciones materialmente incompatibles,

---

[161] David Gauthier ilustra esto con el siguiente ejemplo (en "Economic Man and the Rational Reasoner", cit., p. 107): "las adaptaciones que observamos son respuestas programadas o aprendidas, sin la complejidad implicada por el vocabulario de la creencia, la preferencia y la elección. Una mosca pasa ante el campo de visión de una rana; la pegajosa lengua de la rana sale rápidamente y la mosca es capturada. La rana no cree que una mosca está pasando más de lo que mi termostato cree que la casa está fría cuando responde a una bajada de temperatura activando la calefacción. La rana decide cazar la mosca en la misma medida en que mi termostato decide calentar la casa."

[162] Pues operaciones tales como deliberar, comparar expectativas, deseos, creencias, imágenes de estados de cosas futuros (para elegir entre ellos), etc. sólo son posibles si existe una representación en la conciencia de los actos intencionales de forma que puedan ser tomados como objetos.

[163] En palabras de Gauthier: "Lo que distingue a los seres humanos de otros animales, y proporciona la base para la racionalidad, es la capacidad de representación semántica. A diferencia de un perro, nosotros podemos representarnos un estado de cosas, y considerar si es o no el caso, y si querríamos o no que fuera el caso. Pero al representarlos, los ponemos en relación unos con otros [...] La distancia entre el deseo representado y la acción consciente gobernada por tal representación es pequeña, y es lo único que se necesita para explicar qué significa actuar por una razón." ("Morality, Rational Choice and Semantic Representation", en E. Frankel Paul, *et al.* (eds.), *The New Social Contract*, Oxford, Blackwell, 1988, pp. 173-221; p. 173-174).

entre las que es necesario elegir, y entonces nos vemos compelidos a aplicar algún criterio de elección (o de orden) unificador a nuestras preferencias. En el epígrafe siguiente veremos en qué consiste principalmente este criterio según la Teoría de la Decisión Racional.

Para terminar, recordemos que la concepción instrumental de la racionalidad posee un papel normativo en la teoría del contrato. Dado que se admite que las partes del contrato son agentes instrumentalmente racionales, el pacto depende de que cada uno de ellos lo considere útil para el logro o avance de sus preferencias e intereses. Es decir, la condición del pacto es que cada agente en la posición inicial considere que promueve *sus fines particulares* en la medida suficiente como para que sea racional para él suscribirlo. La "racionalidad del acuerdo" se medirá, pues, por la aceptabilidad individual del mismo.

c) Racionalidad y maximización.-

Decíamos poco antes de concluir el epígrafe anterior, que la capacidad de representación semántica es la condición que posibilita un comportamiento que podamos calificar de racional (al que sean aplicables las categorías de análisis de la Teoría de la Decisión). También identificábamos el carácter instrumental de la deliberación racional: dado un fin, se delibera sobre los mejores medios. Recuérdese el ejemplo de Aristóteles: el objeto de deliberación del médico no es la salud (el fin de su arte), sino los medios adecuados para lograrla en cada caso. Sin embargo, la capacidad de representación semántica y la relación medios-fines, no son suficientes para dar cuenta de la racionalidad práctica. Más bien plantean el problema de la racionalidad práctica misma: como apuntábamos, la representación de distintos estados de cosas impone la necesidad de elegir; se elige conforme a fines subjetivos basados en preferencias, pero éstas no se forman al azar, ni son caprichosas. Distinguimos al agente racional porque establece algún tipo de coherencia a lo largo del tiempo,

de modo que podemos identificar, no sólo sus preferencias presentes, sino también, normalmente, el plan de vida en que se inscriben y los objetivos del mismo[164]; ¿de dónde procede esta coherencia que tomamos habitualmente como dada?

La respuesta ofrecida por Gauthier es simple:

> "Como al representar nuestros deseos tenemos conciencia del conflicto entre ellos, el paso de la representación a la decisión se complica. Debemos, de algún modo, introducir algún tipo de coherencia entre nuestros deseos en conflicto. Y se supone en general que sólo hay un candidato plausible como principio de coherencia —un principio de maximización. Ordenamos nuestros deseos, en relación con la decisión y la acción, de modo que podemos elegir maximizar nuestra expectativa de satisfacción de los deseos. Y al hacer eso, nos mostramos como agentes racionales."[165]

---

[164] De hecho, Gauthier considera que es una necesidad para el agente el ordenar de algún modo coherente nuestros deseos y preferencias. Es una necesidad racional, fuente de la identidad: al unir deseos y creencias en un todo ordenado y mantenerlos así, adquirimos un sentido del yo. El yo se concibe, lejos del humeano haz de percepciones, como un conjunto unificado de deseos y creencias semánticamente representados (cercano a la kantiana unidad de la apercepción). Cfr., p. ej., Gauthier, D., "Morality, Rational Choice and Semantic Representation" (en E. Frankel Paul *et al.* (eds.), *The New Social Contract*, Oxford, Blackwell, 1988, pp. 173-221), p. 219.

[165] Gauthier, D., "Morality, Rational Choice and Semantic Representation", cit., p. 174. Creo que es necesario añadir dos apostillas a este texto, ambas sugeridas por Gauthier. La primera es un comentario que el propio autor incorpora al texto, al agregar que la capacidad de representar nuestros deseos no sólo nos permite hacer prácticamente coherente nuestra conducta, sino además, reflexionar sobre los deseos mismos, cuestionarlos o examinar las relaciones entre ellos. Y esta posibilidad es una dimensión esencial de la racionalidad práctica (que también tendrá su importancia en el argumento contractualista).

La segunda apostilla sólo es insinuada por el autor en una breve nota a pie de página en la que se pregunta si la incorporación de la maximización en la racionalidad práctica no habría de ser defendida, en vez de simplemente supuesta. Desde nuestro punto de vista, esta preocupación es legítima, ya que no es inmediatamente evidente que nuestro comportamiento racional sea siempre maximizador. Sin embargo, en este momento estamos persuadidos de que sí lo es, al menos en el (tal vez un poco trivial) sentido siguiente: todo comportamiento racional puede reducirse a (o explicarse en términos de) maximización. El objetivo de la mayoría de nuestras acciones concretas no es, obviamente, maximizar nada; perseguimos satisfacciones (muchas de ellas imposibles de cuantificar, como la satisfacción del trabajo bien hecho, o las derivadas de las relaciones humanas y los sentimientos) sin apurar los cálculos previos a la selección de una acción.

La maximización ha sido considerada, pues, como la base lógica natural de la decisión[166]. Y la visión de la racionalidad asentada sobre las ideas centrales de instrumentalidad y maximización viene siendo el concepto normativo y explicativo de racionalidad generalmente empleado a lo largo de las últimas décadas por la mayoría de economistas y científicos sociales, el cual ha sido también importado al ámbito de la filosofía moral y política a raíz del desarrollo de la escuela económica de la "Elección Pública".

Se puede decir, por tanto, que el contractualismo liberal abraza la concepción de racionalidad de la que partieran Rawls en la *Teoría de la justicia*[167] o Harsanyi en "Morality and the Theory of Rational Be-

---

Esto podría llevarnos a concluir que nuestro comportamiento es "satisfactor", y no "maximizador". Sin embargo, el principio maximizador se aplica también a la satisfacción: el agente racional busca maximizar aquello que prefiere. Imaginemos un agente que prefiere la satisfacción del trabajo bien hecho al beneficio económico; imaginemos además que es consciente de cómo se realiza un trabajo bien hecho y hasta qué punto merece la pena esforzarse en hacerlo (en relación con la satisfacción que espera). Suponiendo que pudiéramos tener acceso a estas preferencias y, por tanto, estuviéramos seguros de su contenido, diríamos que tal agente no se comportaría racionalmente si buscase el mayor beneficio económico (pues deja de *maximizar* su preferencia); ni tampoco —y esto es más importante— si realiza un trabajo bien hecho, pero con desidia, de modo que la satisfacción final no es tanta como la que podría haber obtenido con algo más de esfuerzo. Si se pudiera admitir que este último caso ejemplifica una acción racional, ello significaría renunciar al principio de maximización, pero ¿cómo sería esto posible? (Obsérvese que no se trata del caso de quien estima que la eventual satisfacción producida por un esfuerzo adicional no compensa dicho esfuerzo y, por tanto, prefiere dedicarse a satisfacer otros objetivos —éste sería un maximizador—; se trata de alguien que, aun siendo consciente de que está en su mano obtener una satisfacción que desea, no la logra por la debilidad de su voluntad). Si, como usualmente se admite, la debilidad de la voluntad no da lugar a acciones racionales, entonces creemos plausible la opinión de que *todo comportamiento racional puede explicarse en términos de maximización*. Coincidimos a grandes rasgos con la idea de Gauthier en "Economic Man and the Rational Reasoner" (p. 108), cuando escribe que "puede que no sea rentable basar cada elección particular en la maximización. Un individuo racional juzgará su procedimiento de decisión en términos maximizadores, y lo abreviará sensiblemente. Puede parecer, así, que en algunas situaciones actúa como un satisfactor, deliberando hasta un cierto umbral y, a partir de ahí, no invirtiendo más energías en la selección. Sin embargo, cuando reflexione sobre tal procedimiento, hallará que tiene una base racional maximizadora".

[166] Cfr. Gauthier, D., "Economic Man and the Rational Reasoner", *cit.*, p. 108. Esta suposición se complementa (Cfr. p. 109) con la confianza en la capacidad *explicativa* del modelo maximizador de la racionalidad práctica.

[167] Cfr. John Rawls, *Teoría de la Justicia*, *cit.*, p. 170. Casi innecesario es añadir que el contractualismo liberal parte de las mismas fuentes de las que lo hiciera Rawls (Cfr. *ibidem*, nota 14).

haviour"[168]. No obstante, habría que matizar esta afirmación, ya que el desarrollo del contractualismo moral modifica la concepción de racionalidad de la que parte al ponerla en relación con la idea de un individuo liberal. En este sentido, Gauthier escribe:

> "El individuo liberal es completamente racional, entendiendo que la racionalidad supone autonomía y capacidad de elegir entre acciones posibles sobre la base de la propia concepción de lo bueno, determinada por las propias preferencias reflexivas. Si la utilidad se define como la medida de la preferencia meditada, el individuo liberal es racional cuando trata de maximizar la utilidad esperada. Aquí enfatizamos la dimensión auto-crítica de la racionalidad práctica, de la que parece carecer la cruda concepción de la racionalidad meramente económica."[169]

Este texto recoge la visión liberal de la racionalidad, que incorpora conceptos ("preferencia reflexiva", "dimensión autocrítica de la racionalidad") ajenos al "concepto de racionalidad (...) que se usa comúnmente en la teoría social"[170], al que nos hemos ceñido hasta ahora. Estos aditamentos nos serán familiares más adelante, pues la teoría moral contractualista se propone mostrar cómo es posible llegar a un concepto (normativo) de racionalidad ligado a las ideas de individuo autónomo y auto-crítico a partir del concepto de racionalidad ya aceptado (y considerado descriptivo o explicativo) por la mayoría de los científicos sociales.

El argumento contractualista adopta como premisa, por tanto, el modelo de racionalidad más débil y comprensivo: la racionalidad como maximización.

---

[168] Publicado por primera vez en 1977, aunque su edición más conocida (por la que lo citaremos) es la incluida en Sen, A., y Williams, B., (eds.) *Utilitarianism and Beyond*, Cambridge, Cambridge U.P., 1982, pp. 39-62.

[169] Gauthier, D., *MA*, p. 346.

[170] Rawls, *Teoría de la justicia, cit.*, p. 170.

Es la propia teoría la que justificará la ampliación o no de este concepto[171].

Queremos insistir en que para una concepción "débil" de la racionalidad, el objeto a maximizar es la satisfacción de preferencias *individuales*. La racionalidad como maximización es incompatible con la moral utilitarista que parte de la existencia de un estado de cosas objetivamente deseable, que todos los agentes racionales deben esforzarse en procurar. El utilitarista moral considera obligatorios los actos que maximizan "la utilidad social", y moralmente reprobables los que no tienen ese efecto. Aplicando una visión maximizadora de la racionalidad, toda acción que no produce un efecto óptimo (comparada con acciones alternativas y factibles del mismo agente) sobre la utilidad social, es moralmente incorrecta, incluso aunque tenga efectos positivos. De este modo, resulta el absurdo de que casi todos nuestros actos serían moralmente malos, pues pueden ser sustituidos por otros netamente más beneficiosos. Kurt Baier, de quien tomamos esta idea[172], pone el anecdótico ejemplo de que muchos actos que satisfacen el deseo inmediato del agente (tales como descansar, o escuchar música) serían considerados reprobables por el moralista utilitario, ya que no optimizan el fin moral (no maximizan la utilidad social).

Esta es una paradoja asociada a la concepción universalista de la racionalidad, que desaparece cuando se toma al individuo como el criterio último sobre los fines de la acción[173]. El acendrado individualismo asociado

---

[171] Aquí debemos anticipar que la teoría moral de Gauthier pretende ser también una teoría de la racionalidad, que amplia y modifica el alcance y posibilidades de la racionalidad como maximización directa. Gauthier sostendrá, por tanto, que su teoría demuestra que la concepción maximizadora, tal como es defendida por los economistas, es incorrecta. Sin embargo, es precisamente la versión más cruda de la maximización la que él acepta como premisa argumental (y, consiguientemente, la que aquí analizamos), por considerar que así su conclusión resulta más plausible.

[172] Cfr. *The moral Point of View*, cit., p. 203.

[173] También parecen desaparecer si se adopta un "utilitarismo de las preferencias", como el postulado por Harsanyi, que parte del principio filosófico de *autonomía de las preferencias* (compartido con el contractualismo). Este utilitarismo refinado de la regla comparte, de hecho, muchas características con el contractualismo. Creo que la diferencia esencial estriba en que el proyecto utilitarista es menos globalmente radical que el contractualismo. Expongo algunas ideas

al enfoque contractualista no elimina, sin embargo, todos los dilemas de la racionalidad. Más bien es fuente de los más contumaces. Pero, desde el punto de vista del teórico del contrato, los problemas causados por el individualismo tienen una solución racional; no así las paradojas del universalismo.

Hablamos, por tanto, de racionalidad individual y de maximización de utilidades individuales que representan preferencias individuales. Desde el punto de vista contractualista, conceptos tales como "el interés general" o la "utilidad social", aceptados por el utilitarismo, carecen, en principio, de sentido.

Debemos reconocer que la atención sobre las preferencias individuales y el expreso reconocimiento de nuestra incapacidad para juzgar sobre su racionalidad (al menos en este nivel del argumento) impone un punto de partida terriblemente estrecho al teórico contractualista. Ya dijimos que el contractualismo liberal descorre el velo de ignorancia rawlsiano[174] y que renuncia a cualquier función de utilidad social; vemos ahora que no admite en principio juicio alguno sobre la racionalidad de las preferencias individuales[175]. A esto hay que añadir que, desde el punto de vista contractualista (que, en este punto, coincide con el de la Teoría de la Decisión), la moral está injustificada en la posición inicial, de modo que las preferencias individuales carecen de contenido

---

deslavazadas en apoyo de esta opinión: el utilitarista parte de que *hay* preferencias morales individuales a las que cabe aplicar criterios maximizadores. El tránsito de las preferencias individuales a la "función de utilidad social" no necesita, por tanto, cuestionarse la deseabilidad e imparcialidad de la sociedad misma. La selección de reglas que definan el comportamiento moral deseable está basada en preferencias individuales *ya morales*. La capacidad motivadora de estas reglas es dudosa. En conjunto, pese a la plausibilidad del intento de Harsanyi, sobre el que habremos de volver, el utilitarismo sigue apareciendo como una justificación racional de una moral social para agentes racionales *que deciden comportarse* moralmente. La raíz racional de esa decisión, a la que el utilitarismo ni se refiere, es el problema del contractualismo moral.

[174] Con lo que se desentiende también del equivalente utilitarista del mismo: el postulado de equiprobabilidad de Harsanyi (Cfr. Harsanyi, J. C., "Morality and the Theory of Rational Behaviour", cit., p. 44 y ss.).

[175] Kurt Baier ha ofrecido una interpretación opuesta de este punto. Cfr. su "Rationality, Value and Preference" (en E. Frankel Paul *et al.* (eds.) *The New Social Contract*, Oxford, Blackwell, 1988, pp. 17-45), p. 28. Volveremos más abajo sobre la misma.

moral[176]. Todo esto configura una "situación inicial" adversa para la moralidad; justamente lo que el contractualismo requiere para su empresa de justificación. Ahora bien, la presunción contra la moralidad ha de estar contrapesada con el empleo del popperiano "método cero", es decir, con una suposición de completa racionalidad; y esto significa, para el contractualista liberal, completa racionalidad de las partes (individuos), expresada también en la completa racionalidad de sus preferencias. Mas volvemos con ello al inicio del círculo, porque hemos establecido nuestra impotencia para juzgar sobre la racionalidad de las preferencias.

En este punto, el equilibrio entre el supuesto de completa racionalidad y el mantenimiento de la razón instrumental maximizadora dentro de sus límites, se logra admitiendo que, si bien no hay un criterio objetivo para valorar los fines del individuo, sí es posible decir algo sobre la coherencia interna de las preferencias individuales. Al fin y al cabo, ése es el fundamento de la Teoría Bayesiana de la Decisión, cuyo axioma de la maximización de la utilidad esperada se asienta únicamente en ciertos requisitos *formales* de los órdenes de preferencias individuales. El contractualista moral añade, a estos requisitos formales, la condición intuitiva de que las preferencias sean "consideradas" o "reflexionadas".

Centrándonos en *MA*, Gauthier dedica una buena parte del capítulo II de la obra al estudio de las preferencias. Parte de la interpretación económica, que supone que un agente prefiere aquello que manifiesta con sus elecciones de hecho y sólo eso; pero para superarla, pues opina que "sea o no este enfoque adecuado para la economía, no es suficiente para la teoría moral"[177]. La teoría moral requiere, según Gauthier, un acceso a las preferencias individuales

[176] Lo cual es, como se ha dicho, una diferencia con el utilitarismo, que parte del reconocimiento de la distinción entre "preferencias personales" y "preferencias morales" (Cfr. arriba, nota 174, y Harsanyi, J.C., art. cit., p. 47).

[177] Gauthier, D., *MA*, p. 27.

alternativo a la elección[178]. Este acceso es posible si se distingue entre dos dimensiones de la preferencia: la dimensión conductual (aquella en la que se fijan los economistas) y la dimensión en cuanto actitud, expresada en el discurso. La coincidencia de ambas dimensiones es un signo de racionalidad; su discordancia, de irracionalidad[179].

La doble dimensión de las preferencias, junto con otras categorías más o menos ambiguas, permite una definición de lo que consideraremos "preferencias meditadas":

> "Las preferencias son meditadas si y sólo si no hay conflicto entre sus dimensiones conductuales y de actitud y se mantienen estables bajo la experiencia y la reflexión."[180]

Las "preferencias meditadas", aproximadamente equivalentes a lo que Harsanyi denominó "verdaderas preferencias"[181], proporcionan el contenido de la utilidad individual (que venimos definiendo como una medida de la preferencia) y, por tanto, el objeto apropiado para la maximización. Además, la dimensión expresa de las mismas autoriza ciertos juicios sobre su racionalidad.

En los casos en que las preferencias expresadas en el discurso y aquellas

---

[178] Ya hemos comentado cómo, al suponer que la utilidad es una medida de las preferencias, y las preferencias son aquello que el agente elige *de hecho*, el economista ofrece una definición *cuasi-* tautológica (y vacía) de utilidad: todo comportamiento, sea el que sea, maximiza la utilidad. Desde este punto de vista, el concepto de utilidad carece de valor normativo. La teoría moral necesita un concepto de utilidad semejante al defendido por los utilitaristas, con cierto (aunque vago) contenido normativo y motivacional. De ahí la necesidad que tiene Gauthier de apartarse de la visión económica de las preferencias, concepto sobre el que se asienta la definición de utilidad.

[179] Aunque con la reserva de que no podemos justificadamente afirmar si lo irracional son los actos (preferencias reveladas) porque no se acomodan a sus actitudes y creencias, o las actitudes (preferencias expresas) porque no expresan sus elecciones de hecho. En este caso, irracionalidad significa simplemente una incoherencia que no podemos localizar con más precisión; una incoherencia que se subsanaría modificando cualquiera de las dos dimensiones (Cfr. Gauthier, D., *MA*, p. 28).

[180] Gauthier, D., *MA*, pp. 32-33.

[181] Cfr. Harsanyi, J.C., art. cit., p. 55.

mostradas en la acción no coinciden, se pueden analizar los procesos o circunstancias que deberían rodear la formación de una preferencia meditada: procesos que no admiten definiciones ni límites precisos, tales como la suficiente reflexión, la información completa o aproximadamente completa, la adecuada experiencia sobre el asunto, etc. La mayoría de los ejemplos en que las preferencias reveladas discrepan de las actitudes se pueden explicar como casos de deficiencia en la información, reflexión o experiencia. Si estos procesos se acercan a la perfección, entonces la actitud y el comportamiento deben coincidir, y hablaremos de una preferencia meditada, medida por la función de utilidad individual. Así, la utilidad tiene un contenido más rico que el que le asignan los economistas[182].

Las ocasiones en que un individuo fracasa en el intento de maximizar la utilidad tal como queda definida o, lo que es lo mismo, las ocasiones en que un individuo actúa en contra de sus preferencias meditadas son ejemplos de debilidad de la voluntad o de falsa conciencia, y pueden ser calificados como conductas "irracionales".

Gauthier insiste, de todas formas, en que el hecho de disponer de este criterio (aunque limitado y ambiguo) para juzgar sobre la racionalidad de las preferencias individuales, no significa en modo alguno abrazar cierto objetivismo axiológico. No hay ningún valor objetivo (ni siquiera el interés del agente) que sirva como medida de lo que es una preferencia racional. El valor a la luz del cual se puede evaluar la racionalidad de las preferencias es subjetivo, viene dado por las actitudes del propio agente, que deben estar incorporadas en las preferencias meditadas. Por eso, pese a la consideración de la dimensión expresa de las preferencias, Gauthier permanece más cerca del modelo económico que del utilitarista[183].

---

[182] Escribe Gauthier: "Estamos ahora en condiciones de añadir algún contenido, aunque todavía insuficiente, al mandato de maximizar. La utilidad es una medida de la preferencia, pero no de la preferencia revelada o de actitud tomadas aisladamente. Es una medida de ambas dimensiones en tanto en cuanto coinciden. Si están en conflicto, sería posible establecer una medida para cada una de ellas, pero no se podrían identificar con la utilidad." (*MA*, p. 28).

[183] Por ejemplo, dado un caso de discordancia entre actitudes y acciones, como el de aquél que, convencido de que fumar es malo para la salud y decidido a promocionar su salud, sin embargo sigue fumando. Dado un caso así, decíamos, podemos denunciar que existe incoherencia entre las

No es eso lo que piensa Baier, quien ha criticado varios aspectos de la teoría gautheriana de las preferencias. Baier formula una crítica sorprendente pues, por un lado, cree que Gauthier intenta separarse del modelo económico sin justificarlo suficientemente; pero por otro lado, le acusa de no ir todo lo lejos que sería necesario en el sentido de conceder prioridad a las actitudes sobre las preferencias reveladas[184]. Da la sensación de que Baier esperara una teoría completa de la utilidad y de la racionalidad deliberativa, olvidando que la teoría de las preferencias está, en la obra de Gauthier, al servicio de una explicación satisfactoria de la elección paramétrica (ampliada luego a contextos estratégicos) que, sin renunciar a la simplicidad de la Teoría de la Decisión supere los límites del enfoque económico permitiendo elaborar una teoría subjetivista del valor.

En su intento de ofrecer un punto de vista no meramente conductual sobre la utilidad y el valor, Gauthier se separa ligeramente de la teoría económica. Tal divergencia se justifica porque una teoría moral no puede aceptar los límites del análisis económico. Pero además, es que dicho análisis obvia la dimensión "expresada" de las preferencias, lo que le impide explicar fenómenos típicamente humanos como la debilidad de la voluntad. Por tanto, consideramos que la pequeña desviación del patrón económico en el análisis de las preferencias está justificada, mientras que una desviación mayor, que condujese a una teoría motivacional sobre las razones para la acción estaría fuera de lugar.

Por otro lado, nos parece que la teoría moral de Gauthier no sufriría excesivamente por el hecho de aceptar la concepción económica de las preferencias. De hecho, a lo largo de la primera parte de la teoría el concepto

---

preferencias expresadas (dejar de fumar) y las reveladas (seguir fumando). Ahora bien, no podemos, desde el punto de vista de Gauthier, afirmar que lo racional sea dejar de fumar (lo es, de hecho, conforme a los valores de este agente concreto). Podría ser que un cambio en las creencias del agente devolviese la coherencia a su acción de fumar. En conclusión, el individuo siempre es soberano para mantener o cambiar sus valores, de modo que el juicio sobre la racionalidad de su acción es siempre relativo y condicional. En este sentido afirmamos que la introducción de las preferencias meditadas no aleja a Gauthier demasiado de la perspectiva económica, ni le acerca al utilitarismo.

[184] Baier, K., art. cit., p. 33.

que se emplea es ése. La matización introducida con la idea de preferencias meditadas parece estar al servicio de una definición del valor (que veremos más abajo) que haga intuitivamente más plausibles los axiomas de la Teoría de la Decisión que valen como principios normativos de la acción racional[185].

d) Principios de decisión racional.-

Como ha quedado ya expuesto, el principio de decisión racional es el de la maximización: un agente actúa racionalmente si y sólo si *maximiza* su utilidad (definida como medida de las preferencias meditadas). Este principio general se modifica según el contexto de la decisión:

Para la decisión individual paramétrica en condiciones de certeza (todas las circunstancias son conocidas y fijas, por lo que cada acción lleva consigo un resultado cierto) el principio se aplica tal como ha sido enunciado[186].

Las decisiones en condiciones de riesgo (cada acción tiene varios resultados *probables* y la probabilidad de éstos es conocida) e incertidumbre (las probabilidades de cada resultado de una acción dada son ignoradas total o parcialmente, de modo que la elección depende de una estimación *subjetiva* de probabilidades) requieren una modificación en el principio de la maximización. Como las acciones no tienen un resultado cierto, no se pueden identificar simplemente con una utilidad, sino que hay que hacer la operación de calcular, según las probabilidades (objetivas o subjetivas) de los distintos resultados, qué utilidad estimada puede asignarse a cada acción. El resultado de ese cálculo se denomina "utilidad esperada" y el principio resultante ordena realizar la acción

[185] Esta opinión viene corroborada por el contenido de las secciones de *MA* que siguen a la discusión sobre las preferencias, dedicadas a las condiciones formales necesarias para que sea posible una elección racional maximizadora en casos de certeza e incertidumbre. Todo el capítulo está enfocado, pues, a hacer inteligible la idea de que la racionalidad consiste en "maximizar la utilidad individual".

[186] Para la maximización en condiciones de certeza las preferencias deben estar en una relación de orden débil ("ser preferido o indiferente a"), que cumpla las condiciones de completud y transitividad (Cfr. Gauthier, D., *MA*, pp. 38-42).

cuya utilidad esperada es mayor, esto es, maximizar la utilidad esperada[187].

La Teoría Bayesiana de la Decisión se ocupa de las condiciones y axiomas de la elección en situaciones de riesgo e incertidumbre. La teoría económica clásica estudió las decisiones en casos de certeza. Ambas constituyen lo que podemos llamar Teoría de la Utilidad.

Mas las situaciones de certeza, riesgo e incertidumbre (elección paramétrica) no agotan las posibilidades de la decisión. Hay un extenso conjunto de acciones en las que los agentes no sólo actúan, sino que *inter*actúan unos con otros, y el resultado de sus acciones depende, eventualmente, no sólo de sus decisiones, sino de las que toman los demás. El complicado problema de decidir y actuar en contextos denominados "estratégicos" (interacción entre dos o más personas) es estudiado por la Teoría de Juegos.

En las situaciones estratégicas presumimos (a) que cada agente es racional en el sentido maximizador expuesto arriba, (b) que cada agente supone que aquellos con quienes interactúa son también racionales. Por tanto, en la interacción racional cada agente trata de maximizar, en principio (al igual que ocurre en la elección paramétrica), su utilidad o su utilidad esperada. Lo que ocurre es que su decisión literalmente *depende* de las decisiones de otros, de forma que una decisión estratégicamente racional debe cumplir estas tres condiciones:

A: La elección de cada persona debe ser una respuesta racional a las

[187] La maximización de la utilidad esperada exige condiciones adicionales a las de completud y transitividad. La razón es que una relación ordinal (x es más preferido que y) basta cuando cada acción está asociada a un único resultado; para elegir basta entonces ordenar "de mayor a menor" las utilidades de esos resultados, sin importar "cuánto más preferido" es un resultado a otro. Sin embargo, el cálculo que ha de determinar la utilidad esperada exige asignar un valor cardinal de utilidad a cada resultado probable (el cual, multiplicado por su probabilidad arrojará la utilidad esperada del mismo). Para que sea posible tal asignación de valores cardinales es necesario que la escala donde se ordenan los resultados por su utilidad permita "medir el intervalo" que separa dos resultados cualesquiera. Las condiciones que permiten una medida del intervalo son la de *monotonicidad* y *continuidad* (Cfr. Gauthier, D., *MA*, p. 44 y 45 para una explicación intuitiva de estas condiciones, así como de los problemas que suscitan).

elecciones que espera que otros hagan.

B: Cada persona debe esperar que las elecciones de todas las demás satisfagan la condición A.

C: Cada persona debe suponer que su elección y expectativas se verán reflejadas en las expectativas de todos los demás[188].

Es fácil ver que estas condiciones no pueden ser cumplidas simultáneamente por dos o más agentes en muchas ocasiones[189]. Típicamente, cuando los intereses son contrapuestos. En esos casos, la maximización de la utilidad es imposible a no ser que se cambie el punto de vista. Esto es lo que hace la teoría de juegos, para explicar cómo es que, de hecho, interactuamos con el afán más o menos logrado de maximizar nuestras utilidades. El nuevo punto de vista consiste en hacer que el objeto de elección maximizadora no sea una acción, sino una *estrategia*. Una estrategia no es una acción, sino un mecanismo o regla para seleccionar acciones[190]. Las estrategias asignan una probabilidad a cada acción[191]. Una vez seleccionada una estrategia, cada actor realizará una de las posibles acciones (con la probabilidad determinada por la estrategia); el producto de las acciones de los jugadores será uno de los posibles resultados del juego[192]. Conociendo la estrategia de cada jugador, podemos calcular el *resultado esperado* de cada combinación de estrategias, multiplicando las probabilidades que las estrategias de los diferentes jugadores asignan a cada acción. Así se forman las matrices que representan los juegos. Éstas son el producto cartesiano de las estrategias puras (aquellas que asignan una

---

[188] Tomamos estas condiciones de *MA*, p. 61.

[189] Para ver un ejemplo, Cfr. Gauthier, D., *Ma*, p. 64.

[190] En los términos de Morton D. Davis: "una estrategia significa un plan de acción completo que describe cuáles serán las acciones de un jugador ante cualquier circunstancia posible" (*Introducción a la teoría de juegos*, Madrid, Alianza, 1986 (4° reimpr.), p. 27.

[191] Se denominan estrategias puras a las que asignan una probabilidad 1 a cierta acción. Las estrategias mixtas asignan probabilidades mayores que cero a varias acciones. Las estrategias mixtas requieren que sea posible definir una medida de intervalo en la escala de preferencias.

[192] Los resultados posibles son varios siempre que las estrategias de los jugadores sean mixtas. Si son todas puras, entonces el producto de cada conjunto de estrategias tiene un único resultado.

probabilidad "1" a determinada acción) de los diferentes jugadores (las estrategias mixtas simplemente asignan distintas probabilidades a cada estrategia pura). La virtualidad del análisis de la teoría de juegos es que permite hacer compatibles las tres condiciones de la racionalidad reproducidas arriba: la elección por cada agente de una estrategia que maximice su utilidad esperada dadas las estrategias que espera que los demás adopten. Si las estrategias de todos los jugadores son maximizadoras, entonces se dice que el juego tiene un resultado en equilibrio. En esta situación puede demostrarse que se cumplen las tres condiciones de la racionalidad[193]. El resultado de equilibrio es difícil si se eligen estrategias puras, pero Nash ha demostrado que en todos los juegos hay al menos un conjunto de estrategias mixtas en equilibrio (es decir, que permiten que *todos* maximicen sus utilidades esperadas dadas las estrategias adoptadas por los demás).

En conclusión, podemos decir que el principio de la racionalidad en casos de interacción prescribe la elección de una estrategia que produzca un resultado en equilibrio.

Apuntaremos los dos inconvenientes de este principio (y,por tanto, los problemas a que se enfrenta la racionalidad estratégica). El primero de ellos es que en muchas situaciones hay no uno, sino varios puntos de equilibrio, y la elección entre ellos carece de un criterio racional definido[194]. El segundo y más importante es que el equilibrio choca frecuentemente con otro principio racional: el principio de optimalidad de Pareto. La optimalidad no tiene que ver con las decisiones estratégicas (cuyo principio de equilibrio ha quedado establecido) sino, en general, con la distribución de utilidades de los distintos resultados. Un resultado es óptimo si y sólo si no hay otro posible resultado alternativo que ofrezca a (al menos) una persona más utilidad sin disminuir la utilidad de nadie. Los resultados óptimos tienen un atractivo inmediato desde el punto de vista de la maximización individual de utilidad. Un resultado no-

[193] La prueba puede verse en Gauthier, D., *MA*, p. 66.

[194] Excepto cuando alguno de los equilibrios es *dominado* por otro, esto es, no resulta más beneficioso para *ninguno* de los jugadores.

óptimo es ineficiente: "desperdicia" utilidad. Por tanto, parece una exigencia razonable que el resultado de la interacción coincida con un óptimo de Pareto. Pero, desgraciadamente, optimalidad y equilibrio (maximización individual de la utilidad esperada) son propiedades diferentes, que muchas veces no coinciden. El ejemplo clásico de esta discordancia es el Dilema del Prisionero, que plantea el problema central de cualquier teoría de la racionalidad estratégica.

Concluimos, por tanto, esta referencia a los principios de la decisión reconociendo la incompatibilidad entre principios igualmente racionales que surge en ciertos casos de interacción; reconociendo, por tanto, los límites de la Teoría de la Decisión Racional[195].

e) La hipótesis de igual racionalidad.-

El núcleo del argumento contractualista de Gauthier es la negociación racional. Ello explica la importancia que el autor concede a la precisión de los requisitos y principios de la racionalidad como maximización. Frente a otros modelos de contractualismo que analizan el pacto social como un acuerdo de contenido necesario, fundado en el consenso unánime y netamente ventajoso para todos en comparación con el estado de naturaleza, el acuerdo moral es analizado por Gauthier como un programa de cooperación negociado que, para los individuos más favorecidos, tal vez sea sólo ligeramente más ventajoso que el estado de libertad natural[196]. Desde ese punto de vista, la negociación racional es el momento crucial del argumento contractualista de Gauthier. Hemos analizado ya cómo se conciben las partes y la racionalidad auto-

[195] Emplazamos, no obstante, al estudio sobre el mercado para mostrar un contexto de interacción ideal en el cual la Teoría de la Decisión Racional podría bastar como guía de la acción, sin exhibir contradicción alguna.

[196] Este es un punto en el que discrepamos de Gauthier, y que ocasiona graves problemas a su teoría. Lo discutiremos más adelante.

interesada que exhiben en la negociación. Consideraremos ahora el supuesto de igual racionalidad, último elemento esencial para un modelo de negociación racional.

La teoría de las negociaciones ha tenido que hacerse cargo de ciertos elementos característicos de los procesos negociadores reales, por ejemplo las amenazas, engaños y "faroles". Desde las negociaciones políticas, sindicales o familiares, hasta los regateos en el zoco, todos nuestros intentos de llegar a un acuerdo desde puntos de partida auto-interesados tienen en cuenta la posible "irracionalidad" de la otra parte: sus descuidos, su carencia de información relevante, su pusilanimidad, su eventual urgencia por concluir el acuerdo, etc. Una teoría de la negociación racional elimina estos factores introduciendo la hipótesis de la igual racionalidad.

Esta hipótesis ha sido criticada por algunos autores[197] por considerarla un supuesto moral subrepticio, que pondría en cuestión el proyecto contractualista de deducir las restricciones morales a partir de premisas estrictamente no-morales.

Gauthier responde a esta crítica en "Morality, Rational Choice and Semantic Representation", donde reconoce "que si la igual racionalidad fuese una demanda moral oculta, o inadmisible por cualquier otro motivo, debería abandonar gran parte del argumento central de *MA*" (p. 186).

En nuestra opinión, la hipótesis de igual racionalidad es criticado más a causa de la propia insistencia de Gauthier en su compromiso para no introducir premisas morales que por alguna implausibilidad intrínseca; pues es

[197] Entre otros, por Gilbert Harman, "Rationality in Agreement", en E. Frankel Paul *et al.* (eds) *The New Social Contract*, Oxford, Blackwell, 1988, pp. 1-16; Albert Calsamiglia, "Un egoísta colectivo: ensayo sobre el individualismo según Gauthier", *Doxa*, 6 (1989), pp. 77-94; y Robert E. Goodin, "Equal Rationality and Initial Endowments", en Gauthier, D. y Sugden, R. (eds.), *Rationality, Justice and the Social Contract*, Ann Arbor, University of Michigan Press, 1994, pp. 116-130. Hay que mencionar también que otros teóricos han apoyado el supuesto, incluso con argumentos diferentes (y tal vez más potentes) que los del propio Gauthier, así entiendo la tesis de Jody S. Kraus y Jules L. Coleman ("Morality and the Theory of Rational Choice", *Ethics* 97 (Julio 1987), pp. 715-749; p. 723) sobre la necesidad de que cualquier teoría moral cuente con criterios pre-teóricos de adecuación para los principios morales. La imparcialidad se puede considerar uno de estos criterios y se expresaría, en un cierto nivel del argumento, en forma de igual racionalidad, sin suponer una petición de principio.

un supuesto incorporado en la tradición contractualista desde Hobbes[198], y nunca antes había suscitado críticas de este tipo. Por otro lado, la creencia de cada agente racional en la racionalidad de aquellos con quienes interactúa (inscrita entre los principios de la racionalidad estratégica) es una convención de la Teoría de la Decisión precisamente por su plausibilidad normativa *prima facie*[199].

Con todo, Gauthier intenta responder a las objeciones mediante dos argumentos ligeramente distintos. Los llamaremos el argumento de la simetría y el argumento del agente.

El argumento de la simetría es como sigue: si las partes en la negociación no fuesen consideradas igualmente racionales, habríamos de admitir que algunos consideran razonable aceptar una distribución *inferior* a la distribución equitativa[200]. Pero entonces, dado que todos son maximizadores de utilidad, sería racional para algún otro reclamar[201] una parte *mayor* que la asignada por la distribución equitativa, ¿sobre qué base podría mantenerse esta asimetría?, ¿no podrían pretender todos el mismo derecho a esa reclamación mayor, lo cual haría imposible el pacto?

---

[198] En el *Leviatán*, parte I cap. 13 leemos: "Porque tal es la naturaleza de los hombres, que aunque puedan reconocer que muchos otros son más ingeniosos, o más elocuentes o más educados; difícilmente creerán que hay muchos tan inteligentes como ellos mismos [...] Pero esto probaría que los hombres son, en este punto, más bien iguales que desiguales. Porque nada es de ordinario mejor signo de la igual distribución de una cosa, que el que todos estén satisfechos con su parte."

Por otro lado, la idea de imparcialidad que el contractualismo hobbesiano incorpora mediante la hipótesis de igual racionalidad cumple la misma función —como ha observado R. Goodin (art. cit. p. 119)— que el "velo de ignorancia" rawlsiano y el "postulado de equiprobabilidad" de Harsanyi. No se trata, por tanto, de un mecanismo ajeno, sino más bien común, al contractualismo.

[199] Gauthier reconoce las implicaciones para la teoría moral de este reconocimiento de racionalidad mutua en "The Incompleat Egoist" (en Gauthier, D., *Moral Dealing*, Ithaca, Cornell U.P., 1990; pp. 234-273), p. 270-271. La igual racionalidad defendida en *MA* y en "Morality, Rational Choice and Semantic Representation" parten, según nuestra lectura, de la concisa idea de "mutua racionalidad" contenida en aquél artículo.

[200] Para esta explicación aceptaremos que *existe* una distribución racional de los beneficios de la cooperación. En su lugar se discutirá cuál es esa distribución, si alguna.

[201] Reclamar, en este caso, significa "no conformarse con menos" o "negarse a entrar en un pacto que le ofreciera menos".

El argumento que hemos llamado "del agente" se fija en la condición de maximizadores de cada uno de los individuos en la negociación. Tal como se ha caracterizado la individualidad y la racionalidad hasta ahora, no hay base alguna para establecer diferencias entre estos individuos. Deteniéndonos en su comportamiento negociador, su mismo afán maximizador les llevará a suponer *ex hipothesis*, que sus semejantes son completamente racionales, que no aceptarán ni más ni menos que lo que ellos mismos aceptarían en su situación. Esta hipótesis elimina los peligros de las amenazas y demás circunstancias distorsionadoras de una negociación. El reconocimiento de la mutua racionalidad es el primer paso hacia cualquier acuerdo, y dado que todos, como individuos maximizadores, están interesados en el acuerdo (pues es globalmente más beneficioso que el punto de no-acuerdo), todos la reconocen. Por último, si la cuota exigida por un negociador es inicua, pensando que los demás se conformarán con una participación menor con tal de lograr un acuerdo (pues un mal acuerdo sigue siendo mejor que ninguno), su argumento se vuelve contra él, pues los demás se negarán a suscribir todo acuerdo no equitativo, y el propio interés del "aprovechado" (para quien también es más beneficioso el pacto equitativo que la situación de no-acuerdo) le conducirá a los términos que lo hagan posible.

Resumiendo, lo que el argumento del agente viene a decir es que, supuesta la perfecta racionalidad de las partes y la ausencia de elementos psicológicos distorsionadores, sólo hay un resultado de la negociación perfectamente racional o equilibrado[202] y, conocido esto, cualquier agente tiene la *misma* fuerza para alcanzarlo, pues su amenaza racional[203] de impedir el acuerdo, siempre surtirá efecto entre agentes racionales.

---

[202] En palabras de Gauthier: "el requisito de la igual racionalidad nos permite seleccionar de los muchos resultados admisibles uno que, al ser igualmente ventajoso para cada negociador, resulta racional para todos aceptar" ("Morality, Rational Choice and Semantic Representation", cit., p. 187-188).

[203] Enfatizamos "amenaza racional" para distinguirla de las amenazas tácticas que se usan en las negociaciones reales. La comprensión de todo el argumento se basa en distinguir claramente que hablamos de una negociación ideal. Las críticas a Gauthier muestran que esta distinción no es siempre fácil.

Ambos argumentos se solapan[204]; entre ambos justifican, a nuestro juicio, la suposición de igual racionalidad, de gran importancia teórica por lo que representa para la concepción de una negociación ideal. Esta suposición permite un gran ahorro de costes de negociación: al reconocerse igualmente racionales, ninguna de las partes del acuerdo pretenderá más que una distribución igual del beneficio, porque saben que es la única factible; así, el resultado del acuerdo es fácilmente predecible por cualquier actor racional. La igual racionalidad garantiza que el resultado de un acuerdo ideal puede ser deducido en cualquier momento (en una situación real), de modo que sirva como criterio efectivo de la justicia de las instituciones concretas[205].

La correcta comprensión de la hipótesis de igual racionalidad y sus implicaciones juega un papel esencial en el argumento contractualista (desde Hobbes, como hemos visto) y debería impedir una crítica relativamente frecuente contra Gauthier, consistente en negar que su teoría moral sirva para individualizar una única estructura social justa, referencia normativa de las instituciones concretas[206]. Según esta crítica, la teoría de la negociación

---

[204] Personalmente, me quedo con el segundo. Creo que el argumento de la simetría contiene una petición de principio (basa su respuesta en la no aceptación de que algunas partes sean "menos racionales", es decir, sigue suponiendo que todas lo son igualmente). El argumento del agente, parte de (a) la perfecta racionalidad de cada agente y (b) de la estructura misma del proceso negociador. Creo que así evita la circularidad.

[205] Cfr. Gauthier, D., "Morality, Rational Choice and Semantic Representation", cit., p. 189. Gauthier llega incluso a comparar el papel de su "negociación ideal" con el que cumpliría la "situación ideal de diálogo" en la ética de Habermas: se trata de un contexto contrafáctico que pondría al descubierto las ventajas injustificadas de algunos miembros de la sociedad.

[206] Quien con mayor claridad ha formulado esta crítica ha sido J.C. Bayón Mohino, *La normatividad del derecho: deber jurídico y razones para la acción*, Madrid, Centro de Estudios Constitucionales, 1992, p. 177. Desafortunadamente, Bayón Mohino se apoya en un argumento de Kraus y Coleman (art. cit., p. 728) que, a nuestro juicio, parte de un análisis incorrecto de las implicaciones que tendría un acuerdo no estrictamente imparcial o equitativo (el único racionalmente factible, según Gauthier). Kraus y Coleman reducen la cuestión al problema de las transferencias improductivas, pues entienden que Gauthier considera que los acuerdos no imparciales se caracterizan por exigir este tipo de transferencias, que sólo serían aceptadas por individuos racionales bajo coacción. Como únicamente los acuerdos equitativos evitan transferencias improductivas, sólo este tipo de acuerdo ofrece una solución racional a la negociación. Kraus y Coleman argumentan que puede haber acuerdos no equitativos que, sin embargo, no exijan transferencias improductivas, es decir, que puedan ser aceptados por individuos racionales. Si esto

serviría para legitimar *cualquier* estructura social aproximadamente beneficiosa para todos, fuese parcial o imparcial. Tal/es estructura/s no podrían arrogarse el privilegio de constituir el criterio de la justicia.

La hipótesis de la igual racionalidad, que es más un supuesto derivado de la lógica de la interacción entre agentes completamente racionales (y completamente interesados, por tanto, en llegar a un acuerdo[207]), asegura el puente entre el auto-interés y la imparcialidad, permitiendo así reclamar la justicia o moralidad (no sólo la racionalidad) del resultado del pacto social.

Con esto concluimos nuestra lectura del concepto inicial de racionalidad. Nos resta únicamente anticipar cómo se entiende el desarrollo de una teoría moral en el marco de la Teoría de la Decisión Racional, asunto al que dedicaremos el epígrafe "g)". Antes de esa explicación, se nos permitirá la interpolación de un excurso que sirva para relacionar el modelo de racionalidad aquí defendido con algunos otros modelos alternativos.

f) Excurso: paradigmas de racionalidad.-

Anticipábamos en el epígrafe "b)" que la dimensión instrumental de la

---

es así, entonces la teoría de la negociación de Gauthier no podría singularizar un resultado, y su papel normativo no podría mantenerse. Pero eso no es así, porque si un acuerdo no equitativo puede mantenerse sin transferencias improductivas aparentes, sólo es a causa de una situación inicial "viciada", que estaría prohibida por una negociación racional ideal, ya que, como veremos en su momento, la misma lógica que impide amenazas u otras distorsiones de la negociación entre agentes completamente racionales y garantiza, así, un resultado equitativo, impide también que determinadas estructuras de la posición inicial (aquellas que implicarían inestabilidad en el acuerdo final) se mantengan en la mesa de negociación y se transmitan al pacto.

[207] Posiblemente, el deseo compartido por toda las partes de llegar a un acuerdo es la implicación más importante contenidoa en la idea de un pacto perfectamente racional entre agentes carentes de "debilidades psicológicas". Este deseo evita las cesiones injustificadas y conducirá invariablemente al acuerdo óptimo, es decir, equitativo. Tal deseo es una hipótesis contra-fáctica (en el mundo real no se observa siempre, más bien al contrario) característica del acuerdo moral que críticos como Bayón Mohino no tuvieron en cuenta o no supieron entender (Cfr. nota anterior). La idea contenida en esta nota procede de T.M. Scanlon, "Contractualism and Utilitarianism" (en A. Sen y B. Williams, *Utilitarianism and Beyond*, cit., pp. 103-128), p. 111.

racionalidad ha despertado escepticismo en cuanto a sus posibilidades filosóficas. El contractualismo liberal la toma como paradigma de la racionalidad individual, y pretende construir un argumento de justificación moral empleándola como única herramienta. Difícilmente se podría esperar que tan improbable programa filosófico recibiera, por parte de la academia, los moralistas o los políticos, algo más que desdén o cierta curiosidad condescendiente. Y, sin embargo, no sólo ha recibido atención, sino más: impulso y crítica, hasta convertirse en el centro del debate filosófico-práctico contemporáneo. Tal vez seamos ingenuos al creer que estos revuelos de gabinete y biblioteca ocurren cuando alguna teoría, casualmente, se parece más al mundo que sus viejas y respetables antecesoras.

Recordemos, a modo de ejemplos y simplificando mucho, las reacciones que, desde distintas tradiciones, ha suscitado el intento contractualista liberal, en especial por lo que se refiere a la concepción de la racionalidad:

La posibilidad de construir una teoría ética de validez general basada en la racionalidad instrumental o estratégica ha sido negada desde la interpretación comunicativa de la racionalidad defendida por Habermas y (en otro sentido) desde la razón "razonable" del último Rawls[208]. Argumentan que la racionalidad estratégica no puede capturar ni desarrollar una verdadera moralidad; la confinan a las interacciones del mercado y la política, mientras suponen que sus versiones de la razón —al menos la habermasiana— son adecuadas para la ética.

Desde una concepción kantiana de la razón práctica se sigue criticando la racionalidad estratégica y el contractualismo con argumentos similares a los que antaño se emplearan contra el utilitarismo.

Pero ello no crea ningún vínculo entre utilitarismo y contractualismo, porque ya hemos dicho que la racionalidad de aquél es universalista y objetivista, lo que se opone al individualismo y subjetivismo de éste. Los utilitaristas

[208] Sobre el debate "racionalidad estratégica vs. racionalidad comunicativa", ver Rubio Carracedo, J., "Los dos paradigmas de la ética: estrategia y comunicación", en *Ética constructiva y autonomía personal*, Madrid, Tecnos, 1992, pp. 59-85.

se quejan de la usurpación de sus ideas fundamentales mientras usurpan ellos las ideas y métodos contractualistas para re-construir un monstruoso utilitarismo de la regla que, de todos modos, no puede escapar a las contradicciones seculares de la escuela.

Por último, el comunitarismo arremete contra la instrumentalidad de la razón liberal, defendiendo que sólo una razón dinámica, teleológica, común y substantiva puede revivir el sentido del bien y la virtud.

Excusamos referirnos a los modelos de razón práctica defendidos por otras tradiciones porque entendemos que —excepto tal vez las teorías feministas (críticas o posmodernas)— tienen poca relevancia para una filosofía práctica[209], o bien pueden ser incluidos en uno de los cuatro grandes paradigmas mencionados: racionalidad comunicativa (post-kantiana), razón práctica kantiana (neo-kantiana), universalista utilitarista y comunitarista.

Buena parte de las concepciones de la ética contemporáneas tienen a la razón instrumental como punto de referencia polémico y se adhieren a uno de los paradigmas de la racionalidad que acabamos de mencionar. Dado que no es posible un comentario mínimamente riguroso de todos ellos, sí queremos al menos exponer brevemente las diferencias principales entre estas interpretaciones y la racionalidad liberal. Al hacerlo no pretendemos exhaustividad ni seguiremos sistema alguno: expondremos las diferencias más relevantes o sugerentes, de acuerdo a nuestra mejor opinión y, distinguidas las concepciones alternativas, intentaremos una defensa intuitiva de la racionalidad estratégica.

La decidida adscripción del proyecto contractualista a la tradición empirista británica lo enfrenta programáticamente con la tradición kantiana. Por ponerlo en términos kantianos, la racionalidad estratégica consiste en —y se refiere a— imperativos hipotéticos y asertóricos inscritos en el gran imperativo

[209] Remitimos, de todas formas, al artículo de Yanis Varoufakis, "Modern and Postmodern Challenges to Game Theory" (*Erkenntnis*, 38, 1993, pp. 371-404), donde puede encontrarse una sugerente alusión a diversas concepciones "alternativas" de la racionalidad, en particular, al anti-concepto posmoderno, a una visión historicista-dialéctica hegeliana y a una reinterpretación de la racionalidad humeana como simple decisionismo.

hipotético de la sagacidad[210]. Desde el punto de vista liberal, el análisis de la racionalidad práctica habría concluido aquí y no se habría extendido en la consideración de los imperativos categóricos[211]. Kant sin embargo, inmerso en un paradigma universalista, no puede concebir un fundamento de la moralidad falible y subjetivo, como sería el resultante de intentar deducir principios obligatorios a partir de los consejos de la sagacidad. Para él, la imperatividad categórica de la moralidad tiene que emanar directamente de una faceta de la racionalidad práctica pura. Por decirlo gráficamente, Kant se prohíbe a sí mismo el camino que el contractualismo liberal abre. Kant considera imposible transitar una vía en la ha de vérselas con el voluble e inaprensible concepto de felicidad; esa es la vía del egoísmo, de los asuntos humanos, no de la moral. La consecuencia es que la fuente de la moral se hurta a la experiencia[212]. La moral surge, así, de la capacidad legisladora de la voluntad, de su capacidad de "poner" normas *a priori* en la experiencia práctica (la acción). Como ha escrito Gauthier (coincidiendo, por cierto, con la magnífica interpretación de J.L. Villacañas[213]):

> "La moralidad, desde este punto de vista, surge de la adscripción a la racionalidad práctica de la misma universalidad encontrada en la racionalidad teorética. Así como la razón teórica descubre leyes descriptivas o explicativas, la razón práctica descubre leyes prescriptivas o justificatorias. El principio maximizador del egoísta es así rechazado por ser insuficiente para la universalidad

---

[210]Cfr. Kant, I., *Fundamentación de la metafísica de las costumbres*, cit., p. 85.

[211] Cfr. Gauthier, D., *MA*, pp. 236-237.

[212] Recordemos el texto de Kant, refiriéndose al imperativo categórico o de la moralidad (*Fundamentación...*, cit, p. 89): "no debe perderse de vista que no existe ningún ejemplo ni forma de decidir empíricamente si hay semejante imperativo, sino que, por el contrario, se debe sospechar siempre que algunos imperativos aparentemente categóricos pueden ser en el fondo hipotéticos..."

[213] Cfr. Villacañas, J.L., "Kant", en Camps, V. (ed.), *Historia de la ética 2. La ética moderna*, Barcelona, Crítica, 1992, pp. 315-404.

inherente a la verdadera racionalidad."[214]

La diferencia entre la concepción kantiana de la racionalidad práctica y la concepción liberal no es la más aparente. Porque aparentemente, la diferencia está en que el liberal (como el utilitarista) acepta sólo la racionalidad de la habilidad y la prudencia, rechaza la idea de una razón pura legisladora, e intenta reconstruir algún tipo de normatividad desde la "débil" base de la racionalidad instrumental. La diferencia es simple, entonces: la tradición empirista intenta encontrar un fundamento para la moral allí donde Kant creyó que era radicalmente imposible hallarlo. Desde el punto de vista kantiano, el liberal está lamentablemente confundido: llama reglas morales a lo que no son más que razonables consejos para una vida feliz. Desde el punto de vista liberal, el kantiano finge unas hipótesis innecesarias para, después de todo, lograr una moralidad ficticia sin capacidad motivadora.

Pero, insistimos, ésta es la diferencia aparente. Porque si el contractualista liberal, en vez de cometer un lamentable error, estuviera en lo cierto, entonces coincidiría con Kant en que la razón posee la capacidad de darse verdaderos mandatos. Diferiría tan solo en el modo en que cree que tales mandatos surgen: la moral no surge —diría el liberal— del principio universalizador de la razón, sino de una necesidad que se impone a cada razón individual en la persecución de su propia felicidad. No hay razón pura; sólo racionalidades individuales. No hay espontaneidad; sólo necesidad racional de establecer un orden y una imagen coherente del yo según el principio de maximización. No hay un mandato categórico; sólo un principio deducido de la misma estructura de la interacción que aparece como medio necesario para la felicidad.

Según nuestra interpretación del contractualismo, la relación en que está dicho principio con las máximas de la acción podría parangonarse con la relación kantiana "ley universal de la razón/máximas de la acción".

Si el liberal está en lo cierto, su visión de la moral no difiere esencial-

[214] Gauthier, D., "The Incompleat Egoist", cit., p. 268.

mente de la de Kant[215]. La diferencia crucial estriba en que el liberal no necesita suponer ni una razón pura, ni su misteriosa potencia como legisladora universal[216].

Pese a sus diferencias, el concepto kantiano de razón coincide con el utilitarista (y con el aristotélico, por cierto) en un punto: todos adoptan una perspectiva objetivista. Esto quiere decir que suponen que la razón tiene la capacidad para decidir no sólo qué es bueno para el agente que razona, sino para cualquier agente en general. El objetivismo no está reñido con el auto-interés: algunas versiones del utilitarismo afirmarían que el mayor bien (o felicidad) para cada uno consiste en perseguir su propio interés, de modo que las acciones que promueven este interés son obligatorias para *cualquier* agente racional. Como sabemos, la racionalidad liberal es subjetivista. Aunque sea posible que los mandatos de una racionalidad objetivista adecuadamente concebida (así como el tipo de acciones prescritas) coincidan en gran parte con las reglas de una racionalidad y una moralidad liberales, su distinto origen no debería dar lugar a confusión: para los utilitaristas hay un fin moral dado (la utilidad social, la utilidad media, la satisfacción de los individuos, etc.) y la razón prescribe los medios necesarios para alcanzarlo; los utilitaristas confían en la capacidad de una razón neutra que decide entre medios alternativos para alcanzar fines dados. Desde el punto de vista liberal, por el contrario, no hay tal razón neutra, con lo que es imposible una elección "objetiva" de los medios —sean acciones, reglas o instituciones— que conducen a los fines morales[217].

---

[215] Aunque su formalismo adopta la forma de procedimentalismo, como muy bien señala Rawls en la *Teoría de la justicia* (México, F.C.E., 1979), p. 294.

[216] Desde el punto de vista liberal, es imposible concebir la razón sin que sea la razón de un individuo (pese a que el argumento del contrato permite una universalización de sus reglas). La idea kantiana de objetividad (retomada por autores como Nagel) es uno de los puntos más alejados del liberalismo.

[217] Esto no quiere decir que el contractualismo liberal no defienda la intersubjetividad de las normas morales post-contractuales y, en cierto sentido, su necesidad racional y su universalidad: el contractualista no es, como defenderemos más abajo, un convencionalista. Si en algún momento pudiera dar esa sensación es sólo porque en este capítulo nos centramos exclusivamente en la concepción (subjetivista) de la racionalidad y no pretendemos tematizar el sentido de una moral liberal. La obligada compartimentación de los temas (y sobre todo el hecho de que ahora hablemos

Como vemos, el individualismo distingue el enfoque liberal del utilitarista. Esta distinción se podría reproducir en el nivel de los fines; pero de ese aspecto ya hablamos antes, en el epígrafe sobre la maximización.

Discutir sobre fines en un contexto utilitarista podría parecer chocante. No en vano el utilitarismo se centra en la dimensión instrumental de la racionalidad, pues considera que la cuestión moral esencial es un problema de cálculo. Sin embargo, el énfasis en los medios no debe hacernos olvidar que, por encima de ellos —tal vez sin apoyo racional o con una justificación empírica contingente— hay siempre un fin común que da sentido a la empresa moral utilitaria.

Interesa destacar la presencia rectora de un fin común en la explicación utilitarista de la razón instrumental, porque es una presencia que la conecta, curiosamente, con el paradigma más opuesto a ella: la idea comunitarista de racionalidad.

Nadie pondrá en duda que el desafío más acuciante de la racionalidad instrumental (en sus versiones liberal o utilitarista) es el planteado por el llamado comunitarismo. Aunque hablar de comunitarismo en general, sin referirse a los distintos autores, es hablar de una posición que no existe, la inevitable generalización nos habrá de servir para decir algo sobre cómo esos autores piensan la racionalidad.

Parten de la que llamaremos "visión dinámica aristotélica" de la razón, que se opone a la prácticamente paralítica razón humeana. La racionalidad dinámica se caracterizaría por incluir fines (los fines no son ajenos a la razón, sino que están incorporados en ella, de modo que un ser racional, por el sólo hecho de serlo, ya está orientado hacia ciertos fines) y por tener una relación directa con la acción, o capacidad motivadora inmediata (de ahí el apelativo de "dinámica"). Según el comunitarista, la razón práctica no es el cálculo que se emplea en la tecnología o la economía, sino aquella cosa, sea lo que sea, que nos mueve a la acción moral, es decir, a hacer el bien o a ejercitar la virtud.

---

de "premisas" teóricas, no de la teoría misma) puede dar lugar a contradicciones aparentes. Esperamos aclararlas más abajo.

La crítica comunitarista al paradigma liberal es demasiado global como para desgajar el aspecto concerniente a la racionalidad e intentar que retenga cierto sentido. El comunitarismo desafía todos los conceptos liberales: individualidad, sociedad, regla, racionalidad instrumental, etc., y su posición sólo se entiende desde la reformulación simultánea de todos ellos.

Por eso, la respuesta liberal ha de ser igualmente global, señalando que el intento comunitarista es una contradicción en los términos, pues tematiza y reivindica un modelo de racionalidad práctica y de vida ética que no es, por definición, auto-reflexivo (al contrario que el modelo liberal moderno). La misma tematización del paradigma denuncia y muestra su pérdida definitiva sin imposible recuperación: reclamar la vuelta a Aristóteles es un *gesto* radicalmente liberal; un gesto de libertad individual nada aristotélico y sí muy moderno. Y el mismo razonamiento puede aplicarse al programa político del comunitarismo más radical. El comunitarismo no es ninguna amenaza a las bases del liberalismo teórico[218], más bien confirma el éxito de la concepción liberal y es un acicate para explicar cómo ha sido posible alcanzar la cohesión y la unidad moral en una sociedad plural, y cómo puede mantenerse en el futuro.

Pero ir más allá en estos comentarios nos alejaría de nuestro propósito de revisar las alternativas a la concepción estratégica de la racionalidad, al que volvemos: nos hemos referido a las concepciones kantiana, utilitarista y comunitarista; nos queda, en último lugar, la razón comunicativa.

Sobre la base de la racionalidad comunicativa se ha construido uno de los argumentos éticos más potentes y válidos para nuestras sociedades liberales. Sus conclusiones normativas no difieren, en muchos aspectos, de las conclusiones del contractualismo liberal más moderado. Metodológicamente, la reconstrucción racional procedimental de la ética es reiterada en ambos casos (lo que apoya la hipótesis de que la racionalidad supuesta debe ser similar), al igual que la idea de pacto unánime, o consenso.

En nuestro concepto, ambos modelos de racionalidad, el comunicativo

[218] Evitamos cualquier referencia a los aspectos políticos, donde la inocuidad del comunitarismo dependerá de qué tipo de creencias se asocien a él.

y el estratégico, expresan dos dimensiones del contenido de la racionalidad tal como es entendida contemporáneamente (aunque cada uno se constituye como paradigma excluyente). Desde nuestro punto de vista, ambos reconocen las varias dimensiones de la racionalidad moderna, pero se distinguen por lo que podemos llamar el orden de prioridad en que ven esas dimensiones. Desde el punto de vista comunicativo, no se niega que un amplísimo campo de la acción humana está dominado por la razón instrumental. Desde el liberalismo no se discute que una comunidad moral debe generar lazos que excedan las relaciones egoístas de mercado, ni que en esos lazos nos reconozcamos más humanos y más racionales que en ningún otro aspecto de nuestras sociedades. Pero si comunicación y mercado, solidaridad y beneficio, comprensión y negociación, discusión e intercambio, conviven en nuestras sociedades liberales, es fácil ver que cada paradigma concede prioridad lógica a uno de los miembros de estos pares, que cree representa la esencia última de la racionalidad humana, y que debería, por tanto, ser el fundamento de una moral racional:

Desde el punto de vista contractualista liberal, el auto-interés, la racionalidad meramente instrumental, es radicalmente anterior. Incluso en términos comunicativos: para entrar en diálogo (sea real o ideal) hay que suponer una lógica instrumental individual que establece las condiciones de posibilidad de la cooperación (al fin y al cabo, el lenguaje es una práctica cooperativa). Por el contrario, desde el punto de vista comunicativo, cualquier transacción (aunque sea auto-interesada) *supone* un acuerdo básico (cuyas condiciones transcendentales indican el fundamento de la ética), pues la transacción implica convenciones, instituciones, prácticas que conformen un marco necesariamente estable que la haga posible, y esa estabilidad depende de mantener un acuerdo esencial que no puede estar fundado en el mero auto-interés[219].

Si pudiéramos generalizar, todas las versiones alternativas de la racionalidad analizadas (excepto, tal vez, la razón comunicativa) caerían dentro

[219] A su vez, el contractualista replicaría que ese acuerdo radical sólo puede suponerse si se parte de que es interesante para cada uno mantener las convenciones que dan paso al intercambio. Así podríamos iniciar un regreso infinito.

de lo que Gauthier denomina "concepción universalista de la razón práctica", que opone a la razón maximizadora instrumental que él sigue. La justificación de su posición frente a la concepción universalista es bien simple:

> "La tarea principal de nuestra teoría moral —generar restricciones morales racionales— es fácilmente lograda por quienes proponen la concepción universalista de la razón práctica. Para ellos, la relación entre razón y moral está clara. Su tarea es defender su concepción de la racionalidad, ya que las concepciones maximizadora y universalista no están en pie de igualdad. La concepción maximizadora posee la virtud, frente a las demás, de la debilidad. Cualquier consideración que proporcione una razón para actuar según la concepción maximizadora, también proporciona tal razón según la concepción universalista. Pero no a la inversa. Según la concepción universalista todas las personas tienen, en efecto, la misma base para la decisión racional —el interés de todos— y esta suposición (la impersonalidad o imparcialidad de la razón) demanda una defensa."[220]

Si se llegase a demostrar lo que Gauthier pide, es decir, que la razón nos permite enunciar fundadamente prescripciones morales universalmente válidas, entonces el intento contractualista sería fútil. El contractualismo viene a remediar, precisamente, la crisis de fundamentos de la moralidad, una crisis debida a la caída de los paradigmas de racionalidad universalista[221]. Al argumentar desde la concepción más débil de racionalidad, el contractualismo intenta asegurarse una plausibilidad que ya no pueden pretender los defensores de una concepción "fuerte" de la razón (que en otro lugar hemos caracterizado como "preñada de moralidad"). La racionalidad estratégica no es un dato

---

[220] Gauthier, D., *MA*, pp. 7-8.

[221] Cfr. Gauthier, D., "Why Contractarianism?", en Vallentyne, P. (ed.) *Contractarianism and Rational Choice*, Nueva York, Cambridge U.P., 1991, pp. 15-30 (en especial pp. 15-16). Hay traducción española de este texto, curiosamente aparecida en 1989: "¿Por qué contractualismo?", en *DOXA*, 6, pp. 19-38.

inmediato: en nuestras interacciones obedecemos reglas de varios órdenes y obedecemos también a nuestros sentimientos. Pero tanto las reglas como los sentimientos varían de sociedad a sociedad, de grupo a grupo, de individuo a individuo. La racionalidad estratégica se descubre como un substrato común a todos los seres racionales, pues está relacionada con la idea básica de que somos seres que, en la medida en que tenemos necesidades y carencias, perseguimos fines. El comportamiento instrumental, y su lógica propia, que la teoría de la elección racional intenta captar, nos hace semejantes, independientemente de las creencias de cada cual; por eso se convierte en la base adecuada para la erección de un marco neutro de convivencia. El contractualismo moral afirmará que ese marco neutro se ordena según principios que captan nuestra idea de la moralidad, pero no es necesario defender esa tesis todavía. Basta recordar que ningún otro concepto de racionalidad puede reclamar un estatuto tan primario y simple como premisa de una teoría moral.

g) La moral como parte de la Teoría de la Decisión Racional.-

Para retomar el curso de nuestra discusión, recordemos que dábamos por concluida, con la alusión a la hipótesis de igual racionalidad, la caracterización de la premisa de la racionalidad, pero dejábamos para este último epígrafe el comentario sobre cómo se encaja una teoría moral en el marco de la Teoría de la Decisión, que es lo que pretende el neo-contractualismo liberal.

En efecto, la originalidad[222] del neo-contractualismo ha sido, desde

[222] Una originalidad que hay que matizar pues, aunque lejos de contar con el refinamiento teórico de la moderna Teoría de la Decisión Racional, la tradición empirista (representada, en este caso, por Hobbes y Hume) inició un tratamiento del ética como una rama de la teoría de la motivación. Para Hobbes, el impulso motivacional básico era el común interés de los hombres en la seguridad; para Hume, la capacidad empática para participar en la felicidad o desgracia ajenas. En cualquier caso, ambos entendieron la ética como la codificación de una parte del aparato motivacional humano. Esta interpretación procede de Nagel; Cfr. *The Possibility of Altruism*, cit., p. 10.

Rawls, confiar hasta tal punto en las herramientas proporcionadas por Teoría de la Decisión que ha llegado a incluir en ella a la propia teoría moral.

> "el mérito de la teoría contractual es que transmite la idea de que se pueden concebir los principios de justicia como principios que serían escogidos por personas racionales [...]. La teoría de la Justicia es una parte, quizá la más significativa, de la teoría de la elección racional."[223]

A modo de apunte, hay que decir que el utilitarismo coincide en esta percepción con el neo-contractualismo. Harsanyi, el representante contemporáneo más cualificado de la escuela, escribía en 1977:

> "Me propongo defender que la aparición de la moderna Teoría de la Decisión ha convertido a la ética en una parte orgánica de la teoría del comportamiento racional."[224]

El análisis de Gauthier ha mostrado, no obstante, que las declaraciones programáticas sobre la conexión entre teoría moral y Teoría de la Decisión Racional no han sido seguidas en toda su radicalidad por sus proponentes. Tanto para Rawls como para Harsanyi, el significado de que la teoría moral sea una parte de la teoría de la racionalidad se reduce a suponer que la ética debe pasar el test de la racionalidad como maximización. Así, para Rawls

> "...los dos principios de la justicia son la solución al problema de elección que plantea la posición original"[225],

es decir, los principios serían aquellos elegidos por agentes racionales en

---

[223] Rawls, J., *Teoría de la justicia*, cit., pp. 33-34.

[224] Harsanyi, J.C., "Morality and the Theory of Rational Behaviour", cit., p. 42.

[225] *Teoría de la Justicia*, cit., p. 144.

determinadas condiciones ideales. La teoría de la elección racional ofrece el modelo (y el principio) para esa elección, pero no deja de ser una herramienta al servicio de una "elección sobre la justicia", cuyo contenido no acaba de incorporarse a la Teoría de la Decisión.

Harsanyi, en un sentido parecido al de Rawls pero, en cierta manera, más clásico, escribe que

> "...los problemas morales deben ser decididos mediante pruebas racionales y el comportamiento moral mismo es una forma de comportamiento racional."[226]

Gauthier ha puesto el dedo en la llaga cuando ha afirmado, sobre este punto, que "Rawls trata los principios de la justicia, no como principios *para* la decisión racional, sino como objeto *de* elección racional"[227]. Frente a esa tibieza, Gauthier se compromete radicalmente (hasta un punto que ni Rawls ni Harsanyi alcanzan) con la afirmación de que la teoría de la justicia es una parte de la teoría de la elección racional. Gauthier trata los principios morales como principios *para* la elección racional en circunstancias de interacción cooperativa. La diferencia deja claro hasta qué punto es íntima la relación entre racionalidad y moralidad para Gauthier:

> "Rawls pregunta: ¿*Qué* elegirían los agentes racionales tras un velo de ignorancia? Responde: elegirían los principios de la justicia. Yo pregunto: ¿*Cómo* elegirían los agentes racionales en la interacción cooperativa? Respondo: elegirían sobre la base de principios morales. Para Rawls, los principios de la justicia constituyen la *solución* a un problema particular de elección racional. Para mí, los principios morales son usados por las personas para *resolver* ciertos problemas de elección racional.

---

[226] art. cit., p. 40.

[227] Gauthier, D., "The Incompleat egoist", cit., p. 236.

> Rawls emplea los principios de la elección racional como herramientas para el desarrollo de su teoría de la justicia. Yo desarrollo la teoría moral como parte de la teoría de la elección racional."[228]

La teoría moral de Gauthier está planteada, en efecto, más como una contribución a la teoría de la elección racional, que como un ejemplo de uso filosófico de la misma. No se plantea —según hace, desde el principio, la teoría de Rawls— como una "búsqueda de los principios para una sociedad justa" (por parte de unos sujetos a los que hay que suponer un "interés por la justicia"), al servicio de la cual se pone no sólo la Teoría de la Decisión, sino todo un esquema heurístico bien complejo, para asegurar unas condiciones determinadas de elección. La teoría de Gauthier, por el contrario, nace como un análisis de las posibles situaciones en que se encuentra un agente racional carente de preocupaciones morales (más bien trata de evitar cualquier restricción a su elección libre), para descubrir que, incluso este tipo de agente encontraría un ámbito de relaciones en que ciertos principios imparciales acordados (que identificamos con la moralidad) representan el modo más racional de actuar. Los principios morales no son fruto de un "interés moral"; son la regla de acción más beneficiosa para cada individuo en contextos cooperativos: son un principio, tal vez el más complejo, para la decisión racional.

Desde esta perspectiva, la premisa de la racionalidad adquiere nuevos tintes, y se entiende mejor el afán de precisión de un autor que no acepta como dados los desarrollos de la teoría, sino que prepara el terreno para ofrecer su propia (y ambiciosa) reformulación.

Gauthier sigue paso a paso un camino riguroso que, aunque puede partir de un egoísmo racional justificado, debe conducirnos fuera del mismo. La

---

[228] *Ibidem*. Mucho más depurada, la tesis se expresa así en *MA* (pp. 2-3): "Desarrollaremos una teoría moral como parte de la teoría de la elección racional. Defenderemos que los principios para realizar elecciones o decisiones entre acciones posibles incluyen algunos que constriñen de un modo imparcial al agente en su búsqueda del propio interés. A éstos los identificaremos como principios morales."

conciencia de la auto-contradicción que conlleva la aplicación directa del principio maximizador en cada acción —el reconocimiento del efecto corruptor que sobre las más simples intuiciones de la racionalidad instrumental tiene la estructura de interacción tipo "Dilema del Prisionero"— debe hacer patente que la correcta interpretación del auto-interés está relacionada con la moralidad; así, el principio moral aparecerá como una parte fundamental de la lógica de la acción racional.

El resultado de esta empresa es, como sabemos, una moral contractual para una sociedad liberal. Así se presenta, al menos, porque gran parte de los críticos —y el propio Gauthier en los años setenta— están convencidos de que el resultado del argumento es sólo un mecanismo pseudo-moral para perfectos hombres económicos, o una "ética del mercado", que en ningún caso se pueden identificar con la moralidad. El Gauthier de *MA* insiste, sin embargo, en el carácter moral de las restricciones acordadas por individuos racionales tal como los hemos caracterizado a lo largo del capítulo. Parece conveniente, entonces, aclarar el concepto de moral que Gauthier toma como "heredado", para hacer más comprensible su tajante afirmación. A esta aclaración dedicamos el punto siguiente.

3. Preconcepción de la moral y teoría del valor

Establecidas las premisas del argumento contractualista, concluiremos el capítulo comentando los conceptos de moral, valor y mercado. Como ya dijimos, estas ideas cumplen distintos papeles en la teoría y requieren un tratamiento diferenciado. El concepto de moral responde simplemente al que Gauthier acepta como "heredado", y su papel en la teoría es relativamente secundario. Sobre el valor, Gauthier desarrolla una teoría subjetivista basada en las ideas de preferencia y utilidad. La función de esta teoría es cualificar al agente racional como maximizador, para diferenciarlo del agente económico. Por último, el mercado merecerá tratamiento aparte, tanto por el peculiar sentido en que es usado en *MA*, como por las críticas y controversias que ha suscitado.

Se nos permitirá que, en vez de comenzar con una definición, lo hagamos hablando de la *función* del concepto de moralidad: Su función es orientadora de una búsqueda. Cuando se busca algo, hay que saber qué se busca; en otro caso nunca estaríamos seguros de haberlo encontrado. Pero la idea previa que tenemos del objeto antes de su hallazgo no tiene por qué ser una idea exacta del mismo (en ocasiones es imposible tal exactitud). El concepto de moral tiene en la teoría contractualista este papel orientador: es una idea general, ampliamente aceptable, de lo que "tradicionalmente" se entiende por "norma moral", o "deber moral", etc. (sin especificar su contenido ni entrar en guerras de escuelas). Este concepto tiene una función secundaria porque no prejuzga los resultados de la teoría: la conclusión del argumento contractualista puede ser un hallazgo que coincida más o menos con la pre-concepción aceptada; o que no coincida en absoluto y, en este último caso, habrá que decidir si es que el resultado de la teoría carece de contenido ético o, por el contrario, se trata de un nuevo contenido que justifica el abandono del

concepto de "normatividad moral" heredado.

En tres lugares de *MA* habla Gauthier sobre el concepto "tradicional" de moralidad. A estos textos añadiremos otro procedente del artículo "The Incompleat Egoist", que aporta un matiz relevante. Reproducimos a continuación los fragmentos literales, para comentarlos acto seguido:

> "Defenderemos la concepción tradicional de la moralidad como una restricción racional a la persecución del interés individual." (*MA*, p. 2)

> "Nuestra teoría debe generar [...] restricciones a la persecución del interés individual que, al ser imparciales, satisfagan la comprensión tradicional de la moralidad." (*MA*, p. 6)

> "La moralidad trata de los agentes, las personas consideradas en tanto que actúan y eligen, y en tanto que implicadas en las consecuencias de sus actos y elecciones." (*MA*, p. 235)

> "La moralidad, tal como la entiendo aquí, proporciona una restricción *interna* al intento directo de que a uno le vaya lo mejor posible." ("The Incompleat Egoist", cit., p. 267.)

La cita de la página 235 de *MA* sitúa la moralidad en el ámbito de la racionalidad práctica. Aunque es una afirmación muy genérica, la referencia a las consecuencias podría entenderse como un sesgo hacia un enfoque de la ética más relacionado con la "responsabilidad" que con la "convicción" o los "deberes absolutos".

Las otras tres citas se complementan. La primera de ellas establece la tesis, la segunda introduce el componente de la imparcialidad, y la última resalta el hecho de que las reglas morales son restricciones internas.

Formalmente, Gauthier pretende enunciar una tesis "neutra", enlazada

con la filosofía práctica que va desde Platón a Kant y que defiende no sólo la capacidad de la razón en su uso práctico, sino su superioridad y dominio sobre los impulsos, deseos e intereses. El conjunto de los tres textos que transcriben esta tesis "tradicional" rechazan por igual una serie de lugares comunes en la ética anglosajona contemporánea, a saber: la concepción humeana de una razón impotente en la práctica (y la ética de la compasión que suele asociarse a ella); el emotivismo, que reduce el papel de la razón en la ética, como mucho, al mero análisis del lenguaje en que ésta se expresa; cualquier posible ética egoísta (aunque el "egoísmo ilustrado" defendido por Baier ejerció una notable influencia sobre la obra de Gauthier); la idea de una ética como expresión de la libertad radical del sujeto, y, finalmente, cualquier postura que defienda el no cognoscitivismo ético. Frente a estas tendencias contemporáneas de la ética, los textos de Gauthier reivindican la concepción "tradicional", apegada a la confianza en la capacidad de la razón —cuya expresión económica y refinada hemos analizado en el punto anterior— en materia práctica.

Como corresponde a un concepto previo, que adquirirá su tinte definitivo al paso que la teoría se vaya desarrollando, Gauthier adopta un compromiso entre las formulaciones modernas de la ética: su concepto puede aplicarse lo mismo a las éticas de corte kantiano que a las utilitaristas o incluso al egoísmo ilustrado (mientras no se entienda como egoísmo ético).

Los matices que introduce (citas segunda y cuarta) le sirven para enfatizar el carácter objetivo (o al menos *intersubjetivo*) que deben tener los principios morales, y su generación autónoma[229].

Ambos aspectos son especialmente necesarios desde el punto de vista del contractualismo moral[230]. El primero porque se tiende a ver el contractualismo como un mero convencionalismo, en que el interés particular sigue

---

[229] Al comentar, respectivamente, que las restricciones al interés individual deben ser *imparciales* e *internas*.

[230] Así fue reconocido por Gauthier en "Morality, Rational Choice and Semantic Representation" (cit., p. 177), donde, al responder a sus críticos, pone el énfasis sólo en los dos rasgos que comentamos: "Afirmaré que una restricción moral se caracteriza por dos rasgos: es interna, pues opera a través de la voluntad o capacidad de decisión del agente, y opera de tal modo que satisfaga cierto criterio de imparcialidad entre personas. Esto basta para mi propósito reconstructivo."

primando incluso cuando se trata de obedecer las reglas acordadas, de modo que pacto, reglas, obediencia y ética en su conjunto tendrían una base inevitablemente subjetivista. La idea de imparcialidad elimina las concepciones de la justicia de corte tiránico, como la defendida por Trasímaco en el libro I de *La República*, y —mediante la idea de un acuerdo perfectamente racional— también los convencionalismos, como el expuesto, en el libro II, por Glaucón.

El segundo matiz, referido al carácter interno de las restricciones morales, no es sólo una pincelada kantiana, sino que tiende a asegurar que no se aceptará como pacto moral un pacto de tipo hobbesiano, que, si efectivamente restringe la persecución individual del interés en aras del beneficio común, lo hace sólo gracias al soberano, dando por supuesto que no existe la posibilidad de poner un límite interno a la lógica auto-interesada de la racionalidad individual. Para que el resultado del pacto pueda considerarse moral hay que exigir que las restricciones que funda sean internas, esto es, puedan mantenerse sin necesidad de incorporar una garantía coactiva.

Los rasgos que Gauthier asocia a un concepto tradicional "estándar" de moralidad son, ciertamente, lugares comunes: las reglas morales se oponen al interés privado y lo limitan (porque conciernen al interés común o general); son racionales, luego la empresa de su justificación filosófica es factible; son imparciales, es decir, objetivas o independientes de la voluntad del agente y, finalmente, se trata de normas autónomas o internas, frente a la heteronomía de los usos sociales y las leyes jurídicas y políticas. Pero si estos trazos son comúnmente usados, no lo es tanto la formulación concreta elegida por Gauthier. Nosotros creemos hallar una explicación para la misma en la influencia ejercida por la concepción de la moral expresada por Baier en *The Moral Point of View*, una obra que, en ciertos aspectos, pre-figura y alimenta el proyecto filosófico de *MA*.

Baier proporciona, a la vez, una de las ideas nucleares de la teoría de Gauthier y buena parte de su léxico característico, cuando escribe:

"La verdadera razón de ser de una moralidad es aportar razones

> que se impongan a las razones de auto-interés en aquellos casos en que sería perjudicial para todos que cada uno siguiera su auto-interés."[231]

El fundamento que Gauthier encontrará para la restricción que la moral implica coincide con el que Baier reconoce en el texto reproducido[232] y, por tanto, no debe extrañar que la concepción inicial de la moralidad se exprese en términos de "restricción al auto-interés".

Aparte de esta huella fundamental en el modo de concebir y describir el papel de la moral, la obra de Baier deja su sello sobre dos convicciones gauthierianas que configuran su mirada inicial hacia la moralidad: se trata, por un lado, de la certeza de que hay una moral racional y, por otro, la tesis de que las restricciones al auto-interés no tendrían sentido para un ser solitario, es decir, que la moral es un fenómeno necesariamente social. Esta segunda tesis no está suficientemente reflejada en los textos que seleccionamos (excepto, parcialmente, en el tercero), pero es uno de los puntos de partida de una teoría contractual de la moral.

La certeza de que puede haber una moral racional está relacionada con el cognoscitivismo ético del que Baier había partido[233]. La idea de una moral racional no puede negar, sin embargo, que nuestras moralidades concretas se alejan frecuentemente de su criterio. Ahora bien, en cuanto a qué moral sería aceptada por agentes perfectamente racionales, Gauthier conviene con Baier en su tesis fundamental[234]. La moral racional expresa —para ambos autores— no tanto unos intereses ya morales, como los intereses "verdaderamente racionales" de cada agente. Gauthier podría muy bien suscribir el siguiente texto de

---

[231] Baier, K., *The Moral Point of View*, Ithaca, Cornell U.P., 1958, p. 309.

[232] Cfr. también p. 314.

[233] Cfr. Baier, *op. cit.*, p. 179-180.

[234] Cfr. "The Incompleat Egoist", cit., p. 254.

Baier:

> "Ya no hay ningún misterio en por qué somos morales —aquellos de nosotros que lo somos. Porque ser moral es simplemente un caso especial de seguir a la razón..."[235]

La segunda tesis de Baier que habría influido en Gauthier era la referida a las condiciones de la moralidad: la moral no tiene sentido para un ser humano solitario en una isla desierta[236]; la moral sólo tiene sentido en sociedad porque su fundamento racional (restringir el auto-interés cuando es mutuamente beneficioso hacerlo) sólo puede darse en interacción. ¿Qué justificación racional podría darse para la acción por la que un agente solitario dejase de hacer lo más conveniente y beneficioso para sí mismo? Antes de cualquier calificación moral, tal acción sería considerada simplemente irracional.

Esta concepción (utilitaria) de la moralidad destierra las consideraciones morales de aquellos ámbitos en los que la interacción cooperativa no se da (isla desierta), o no puede darse (estado de naturaleza), o carece de efecto práctico. Este último caso sería el del mercado perfectamente competitivo (del que hablaremos en el punto siguiente) donde, a pesar de la interacción entre agentes, las decisiones de cada uno tienen una influencia despreciable en cómo les va a los demás, por lo que todos pueden decidir *como si* se hallaran en una isla desierta, en medio de unas condiciones naturales fijas.

Lo decisivo de la idea que Gauthier toma de Baier es que en los contextos de decisión no estratégica, las restricciones morales carecen de justificación racional. Esto tiene dos consecuencias: la primera es que la moral que justifica Gauthier tendrá la apariencia de una ficción instrumental (que la última parte de *MA* intenta desmentir). La segunda es la apertura al modelo de

---

[235] Baier, *op. cit.*, p. 298.

[236] En las afortunadas palabras de Baier (que repetirá Gauthier): "si los individuos vivieran por su cuenta y no pudieran afectarse unos a otros, entones, moralmente hablando, no habría nada que no pudieran hacer o de lo que tuvieran que abstenerse. Un mundo de Robinsones Crusoe no tendría ninguna necesidad de una moralidad, ni hallaría utilidad alguna en ella. Las distinciones morales no se aplicarían allí" (*op. cit.*, p. 215).

justificación contractualista, ya que la hipótesis de contextos en que las distinciones morales (pero no las racionales auto-interesadas) carecen de relevancia, invita a la reconstrucción racional de la moralidad a partir de un estado de naturaleza concebido como un contexto pre-moral[237].

En definitiva, nos parece evidente que la obra de Baier es la piedra de toque para interpretar la pre-concepción de la moralidad en *MA*. La definición de la moral como una restricción al auto-interés cuando beneficia a todos por igual el que todos se abstengan de perseguirlo, late en los textos de Gauthier (a veces expresamente); y, por otro lado, el proyecto filosófico consistente en aceptar el egoísmo racional para transcenderlo, justificando su restricción por razones asimismo auto-interesadas (el beneficio colectivo) es, casi literalmente, el mismo que intenta Gauthier, con la única diferencia del método. El contractualismo justifica a un nivel más profundo lo que Baier había simplemente enunciado.

La moral queda concebida, por tanto, como un conjunto de restricciones internas que, si han de estar racionalmente justificadas, tienen que basarse en el mismo interés que restringen. Se trata de una concepción *normativista* de la moral, en la que cualquier alusión al bien o al valor brilla por su ausencia. Y no es que el valor deje de jugar un papel importante en la concepción contractualista de la racionalidad práctica[238]; todo lo contrario: la teoría del valor abrazada por el contractualismo liberal permite escapar de los estrechos márgenes de una racionalidad puramente económica, para conformar una

---

[237] Sobre esto, Baier mantiene una tesis empíricamente cuestionable que haría plausible un contractualismo entendido como reconstrucción histórica (y no sólo racional) de la moralidad: "..la moralidad es un sistema de reglas comparativamente sofisticado, y tenemos que admitir la posibilidad de sociedades no-morales o pre-morales, igual que hay sociedades no-políticas o pre-políticas" (*op. cit.*, p. 179-180).

[238] Aunque no debe suponerse que este papel supone una "moralización" de la razón, o una identificación de la moralidad con la racionalidad, como ha supuesto Jean Hampton en su interpretación de Hobbes: Cfr. *Hobbes and the Social Contract Tradition* (Cambridge, Cambridge U.P., 1986), p. 28 y ss. Aunque debatiremos este aspecto en el capítulo siguiente, queremos dejar constancia de que la teoría liberal del valor no constituye, por sí misma, "una teoría ética". Es sólo uno —el más contingente, por cierto— de sus componentes.

concepción global (y de mayor alcance) de la racionalidad práctica. El relativo "poco peso" de esa teoría en la ética liberal se debe a que el valor es concebido como relativo y subjetivo. El papel que en una ética "de bienes" jugarían los valores objetivos es reemplazado aquí por el auto-interés, expresión que encierra una comprensión de los valores inspirada lejanamente en la rudimentaria concepción hobbesiana del bien. El auto-interés determina la configuración del valor y, simultáneamente, es el fundamento del acuerdo normativo básico que permite la realización individual del mismo.

Desde el punto de vista liberal, comprometido, como hemos visto, con el individualismo metodológico y con la concepción instrumental de la racionalidad, el valor sólo puede concebirse como subjetivo y relativo. La posibilidad y carácter de una normatividad moral racional debe aceptar el valor, así concebido, como dado. Sin embargo, desde el punto de vista del análisis, la concepción liberal del valor requiere, al menos, un breve comentario:

En primer lugar, precisaremos qué entiende Gauthier por "subjetivo" y "relativo" y, después, intentaremos defender el punto de vista liberal frente a otras concepciones del valor.

El análisis del valor emprendido por Gauthier toma como base la teoría de la elección racional ya explicada en el punto anterior y el marco de racionalidad práctica en que se inscribe. Desde el punto de vista de la acción, el valor tiene un papel explicativo: el valor asignado a ciertos estados de cosas hace inteligible la elección. Sin embargo, esto es decir bien poco si no se añade algo sobre la naturaleza del valor. Esta naturaleza estará, de acuerdo con la concepción mínima de la racionalidad individual, relacionada con las preferencias subjetivas. Pero si se identifica el valor con las preferencias de primer grado, o inmediatas, es decir, con lo que el individuo desea en cada momento, no hemos avanzado en la comprensión y crítica de la elección. Es necesario, como vimos en el punto *2. c)* de este capítulo, precisar convenientemente el concepto de "preferencias meditadas", que permite una definición de utilidad de mayor alcance que la económica. Si la utilidad se define como una medida

de la satisfacción de las preferencias coherentes y meditadas, entonces puede —según Gauthier— identificarse con el valor:

> "Si se cumplen las condiciones de coherencia en las preferencias, podemos introducir una medida de los objetos de preferencia. Si se cumplen también las condiciones para una preferencia meditada, podemos identificar esta medida con el valor, y su maximización con la racionalidad."[239]

De este modo se acomoda el concepto de valor en la descripción de la racionalidad. En la medida en que las condiciones para la medida de la preferencia y las condiciones para una preferencia meditada no conciernen (como veíamos también en el punto anterior) al contenido de lo preferido, sino que son puramente formales, el hecho de que se cumplan no implica que los valores de todos los agentes racionales sean o deban ser semejantes. Las preferencias son individuales, y determinarán los valores de cada individuo con independencia de los valores de los demás.

La asimilación del valor a la utilidad es un expediente simple para armonizar la herencia hobbesiana, el sencillo análisis económico y las complejas implicaciones filosóficas de la teoría del valor. Esta identificación exime a Gauthier de una defensa larga y polémica del punto de vista liberal sobre el valor y, por otro lado, es enormemente precisa en cuanto a la descripción de dicho punto de vista.

En efecto, al equiparar valor y utilidad, queda claro el carácter subjetivo y relativo que asignan al primero los teóricos liberales. Veamos por qué.

El valor es una medida de la preferencia. Y una medida está en función de aquello que mide: allí donde no hay preferencia, no hay valor. Un objeto posee valor sólo en cuanto está en una relación particular con uno o varios agentes: es preferido. Aunque en el lenguaje común se adscriben valores a los objetos mismos, ello puede interpretarse como una generalización hipotética:

---

[239] Gauthier, D., *MA*, p. 24.

como los objetos de tal tipo suelen ser mayoritariamente preferidos, cabe inferir que si cierto agente (o el "elector medio") tuviera la oportunidad, preferiría tal objeto con tal intensidad (o le asignaría tal valor). Pero esta generalización no separa al valor de las personas y sus actividades; el valor no es una característica inherente a las cosas ni algo que exista fuera de las relaciones afectivas entre los agentes y sus objetos de preferencia. Y concebir así el valor, es concebirlo como *subjetivo*.

Así pues, afirmar que el valor es subjetivo significa afirmar que su fundamento está en la actitud de los sujetos hacia los objetos. Es importante resaltar este punto para evitar la confusión de identificar una doctrina subjetivista del valor con aquella que considera que sólo los estados de cosas *referidos al sujeto* poseen valor intrínseco. El subjetivismo de los valores no limita en absoluto el contenido de los objetos valorados.

La concepción subjetivista del valor no implica, por otro lado, ni que los valores sean arbitrarios, ni que sean incognoscibles. Tal como se ha definido la utilidad en relación a las preferencias conductuales y de actitud, es evidente que los valores no son meros nombres asignados a la primera ocurrencia caprichosa de los deseos del agente. Más bien son actitudes plenamente consideradas hacia ciertos estados de cosas, dadas las creencias sobre los mismos. Ciertamente, no tenemos criterio racional substantivo alguno para juzgar el nexo entre las creencias y las preferencias meditadas, pero eso no significa que éstas sean arbitrarias.

Sobre la cognoscibilidad de los valores, Gauthier argumenta que el subjetivismo no significa que los valores sean incognoscibles. Lo que el subjetivismo sí niega es que el conocimiento de los valores sea otro que el empírico ordinario. Según la concepción subjetivista, el conocimiento de los valores trata de los afectos y de los procesos (cognitivos) de evaluación; no trata de ningún misterioso ámbito axiológico aprehensible sólo mediante cierta forma de intuición distinta de la experiencia sensible.

En cuanto al carácter *relativo* del valor, es una afirmación independiente de su calificación como subjetivo, aunque tenga que ver con ella. Podría suponerse que, aunque el valor fuese determinado por las preferencias

(afecciones) subjetivas, éstas a su vez podrían estar determinadas por una regla absoluta o universal, de modo que pudiera hablarse de un único concepto (o punto de vista) del bien, del que los valores subjetivos participan o al que obedecen. No es esta la explicación liberal. Los valores dependen de las relaciones afectivas de *cada individuo*, y la determinación de éstas es autónoma. La estimación dependerá del punto de vista individual desde el que se realiza, sin que sea posible concebir un "punto de vista absoluto". El individualismo de partida excluye, por otra parte, relativismos no individuales, que relacionarían el valor con el bien para una familia, grupo o clase social; y, por supuesto, excluye la concepción absolutista que vería el valor para cada individuo como la participación en un "bien" universal.

El absolutismo en la concepción de los valores implicaría que el *punto de vista* desde el que se evalúa no tiene consecuencias sobre la evaluación misma; el relativismo sostiene lo contrario: el punto de vista del agente que evalúa *determina* el valor. No hay "razones objetivas" que hagan racional para todos tener cierto deseo o promover cierto estado de cosas. Es decir, no hay razón alguna para creer que lo que es valioso para un agente ha de serlo para todos los demás[240].

A pesar de que se suelen presentar unidos, no hay una relación necesaria entre relativismo y subjetivismo. Gauthier pone el ejemplo de Hobbes, quien, según cierta interpretación, podría ser considerado relativista y, a la vez, objetivista[241].

---

[240] Obsérvese que el hecho de sostener un absolutismo o universalismo de corte subjetivista o egoísta *parece* contradecir esta afirmación. Tal universalismo axiológico afirmaría, por ejemplo, que es valioso para cualquier individuo defender sus intereses (*su* país, *su* comunidad, *su* bienestar, etc.). Parece que el bien es, según esta concepción, distinto para cada agente, pero no es así: el bien se identifica con un objeto generalizable (el interés subjetivo). Quienes propugnan esta tesis (que podemos considerar defensores del "egoísmo ético") son absolutistas; supondrían que los "intereses de cada agente" podrían ser valorados *desde cualquier punto de vista* suficientemente informado. Frente a este concreto modelo de absolutismo, el relativismo liberal resalta su carácter autónomo.

[241] Según esta interpretación (probablemente errónea, apunta Gauthier en *MA*, p. 51), la auto-preservación se podría considerar un bien objetivo: proporciona una norma para nuestras preferencias. Cfr. también nota anterior.

Por otro lado, el subjetivismo y el relativismo de los valores no impiden que en el lenguaje común se empleen expresiones que asignan valores absolutos a los objetos o a los estados de cosas. Esta costumbre puede explicarse como una generalización de las preferencias que muchos individuos (o algunos individuos relevantes) muestran hacia los estados de cosas en que participa un cierto objeto. A base de generalizaciones puede llegar a asignarse a un objeto cierto valor positivo o negativo. Pero, además, hay un mecanismo para comparar gustos y preferencias. No es un mecanismo perfecto, pero se usa con bastante éxito: el intercambio de mercado. Según lo preferidos que los objetos son, el intercambio de mercado les asigna valores compartidos por todos los que participan en el mismo. Así, si yo prefiero manzanas verdes a manzanas rojas, pero sé que las rojas "valen más" (en el sentido de mercado), y se me da a elegir entre un kilo de las primeras y un kilo de las segundas, elegiré éstas últimas (las "valoraré más", paradójicamente), porque conozco que el kilo de manzanas rojas *equivale* a (o puede ser intercambiado por), digamos, kilo y medio de manzanas verdes. Así, al elegir conforme al valor de mercado maximizo mi preferencia.

No se debe olvidar que el valor intersubjetivo de cambio sigue dependiendo de las preferencias subjetivas (de hecho, es proverbial cómo las bruscas mutaciones en las preferencias trastocan las valoraciones del mercado). Este ejemplo muestra que el hecho de acudir en ocasiones a criterios "objetivos" de valoración no refuta el subjetivismo.

Un intento más serio de refutación del subjetivismo proviene de la afirmación de que los valores objetivos *explican* nuestros deseos e inclinaciones y, en última instancia, nuestras decisiones. Tal afirmación se apoya, primero, en la idea de que las preferencias han de tener algún motivo o causa y, segundo, en el expreso reconocimiento, por parte de los agentes, de su convicción de que prefieren los objetos o estados de cosas que de hecho prefieren debido al valor intrínsecos de los mismos. En cuanto al papel explicativo de los valores objetivos, es fácilmente desenmascarado si se profundiza sobre cuál es la *mejor explicación* de nuestras observaciones[242]. El análisis de nuestras

[242] Cfr. Gauthier, D., *MA*, p. 56.

acciones y elecciones muestra que la mejor explicación que podemos dar de las mismas es el esquema que comentábamos en el punto anterior: la elección maximiza la satisfacción de las preferencias dadas las creencias. El valor objetivo no juega papel alguno en tal explicación. Allí donde el único modo de explicar nuestras observaciones empíricas es introducir la hipótesis de que los objetos poseen propiedades intrínsecas (como ocurre frecuentemente en las ciencias), estas hipótesis se aceptan porque forman parte de la mejor explicación posible de lo observado. Pero en el caso de la acción y la elección no hay tal necesidad, por lo cual el valor objetivo es una inútil redundancia explicativa y, como dice Gauthier, "debe ser eliminado de la faz del universo por la navaja de Ockham".

El segundo apoyo de la defensa del objetivismo se refería al hecho de que la información expresa que muchos agentes ofrecen sobre sus acciones incluye referencias a valores objetivos o absolutos. A esto podría añadirse que muchas acciones sólo pueden explicarse si se supone *la creencia* del agente en la existencia de valores absolutos. Pues bien, tanto la idea que los agentes tienen de los motivos de su acción, como la explicación psicológica de los mismos, son datos empíricos cuya constatación no nos obliga a creer en lo que afirman[243].

En definitiva, la teoría del valor defendida por Gauthier sostiene que "lo que es bueno es bueno en última instancia porque es preferido, y es bueno desde el punto de vista de aquellos, y sólo aquellos, que lo prefieren"[244]. Si bien la teoría no es desarrollada completamente, al menos establece el que podemos llamar "marco axiológico" de una moral contractual. Como corresponde a una ética liberal, el compromiso consiste en demostrar la racionalidad de restricciones normativas morales universales dado que el valor es subjetivo y relativo.

---

[243] Gauthier hace un clarificador paralelismo entre la creencia en valores objetivos y la creencia religiosa: el hecho de que una parte del comportamiento de muchas personas sólo pueda explicarse suponiendo sus creencias religiosas, no nos compromete a creer nosotros mismos (Cfr. *MA*, p. 58).

[244] *MA*, p. 59.

4. El papel del mercado

Tras el planteamiento de las premisas teóricas y del marco axiológicamente plural en que se inscribe una moral por acuerdo, el siguiente paso debería ser la identificación del esquema de interacción racional de los individuos en el estado de naturaleza o posición inicial. Los postulados teóricos del contractualismo deberían concluir en este punto, para dejar paso a la deducción de los problemas generados en el estado de naturaleza y la defensa de su posible solución. Sin embargo, antes de iniciar el argumento contractualista propiamente dicho, Gauthier dedica un capítulo al mercado perfectamente competitivo[245], una institución fantasmagórica cuya descripción no ha convencido a la mayoría de los críticos y cuyo papel en la teoría resulta dudoso.

En efecto, hay una clara línea argumentativa que discurre por los siguientes pasos: (1) identificación de las condiciones y principios de la racionalidad individual, que constituyen el elemento teórico básico para definir el estado de naturaleza como un conjunto de individuos perfectamente racionales; (2) descubrimiento de la auto-frustración de la racionalidad individual (imposibilidad de alcanzar al mismo tiempo maximización individual y optimización colectiva) expresada en las situaciones del tipo del Dilema del Prisionero, que sirve de modelo de "interacción natural"; (3) justificación de la cooperación (para armonizar de nuevo la racionalidad individual y colectiva), que se implementa mediante la negociación y el contrato, únicos mecanismos adecuados para producir principios de cooperación que logren dicha armonización respetando escrupulosamente los intereses individuales, (4) explicitación de las condiciones necesarias para el cumplimiento de lo pactado entre maximizadores racionales, y (5) defensa de que los principios de cooperación

[245] Un capítulo inspirado en el artículo de 1982 "No Need for Morality: The Case of Competitive Market", en *Philosophical Exchange*, nº 3, pp. 40-54.

acordados son principios morales. En esta línea argumentativa el mercado brilla por su ausencia. A lo sumo, el mercado podría debatirse como una de las instituciones legitimadas por los principios de cooperación racional, es decir, como una de las instituciones de una sociedad liberal justa; pero ¿cumple algún papel en el estadio inicial del argumento?

Nosotros creemos que sí: para describirlo con una palabra, diríamos que el papel del mercado es "metafórico". El mercado aparece en este nivel del argumento como metáfora de un inexistente reino de libertad absoluta, donde agentes perfectamente racionales y completamente informados actuarían de modo auto-interesado para dar lugar a un resultado mutuamente beneficioso y socialmente óptimo sin necesidad de restricción moral alguna. Por decirlo así, el mercado ideal es un "anti-estado de naturaleza", porque, al igual que la hipótesis del estado de naturaleza conduce al pacto social y lo justifica, la situación del mercado perfecto hace innecesario e injustificado cualquier pacto restrictivo. Frente a las condiciones del mercado perfecto, las características del estado de naturaleza, y la consiguiente necesidad de la moralidad, aparecerán con más claridad. No negamos que tal vez es una metáfora mal elegida —sin duda lo es, pues ha sido muy mal comprendida— porque todo mercado, incluso el mercado ideal perfectamente competitivo, es ya una institución moral, donde han de respetarse varios límites externos a la acción[246]. No obstante, queremos vindicar en este epígrafe el papel del mercado en una teoría contractual de la moralidad, aunque para ello debamos apartarnos de la opinión del propio autor de *MA*[247] y modificar algunos aspectos de su formulación.

Desde luego, el papel "metafórico" del mercado en este nivel del argumento *no es central*[248]: no interviene en la línea argumentativa principal

---

[246] Como reconoce Gauthier en *MA*, p. 85.

[247] En efecto, David Gauthier ha venido a reconocer, tras las muchas críticas recibidas por su descripción del mercado, que se trata de una hipótesis prescindible y que, en todo caso, debería ocupar otro lugar en la obra.

[248] En contra de lo que expresamente afirma Gauthier en *MA*, p. 84: "el mercado es un interés central de nuestro estudio". Entendemos que esta afirmación ha dado lugar a grandes malentendidos entre los críticos. No obstante, el mercado sí es un aspecto central en la configuración de la

de la moral por acuerdo. La triple función que cumple concierne al fortalecimiento de las premisas argumentales, a la precisión del estatuto ontológico de una moral por acuerdo y a la enfatización de las condiciones de la interacción natural que hacen necesario y racional el contrato moral. Estas precisiones coadyuvan al éxito y claridad del argumento, pero no le son esenciales. Las funciones del mercado se basan en que la construcción heurística del modelo de competencia perfecta permite pensar una estructura de interacción tal que las acciones y decisiones de un conjunto de agentes racionales maximizadores que siguen su auto-interés sin restricción alguna, arrojan un resultado individualmente maximizador y socialmente óptimo[249]. La *posibilidad* de pensar dicha estructura de interacción garantiza que nuestras premisas no incluyen supuestos morales. El mercado perfectamente competitivo confirma que las hipótesis metodológicas aceptadas permiten conformar una estructura de interacción óptima (colectivamente racional) a partir de las racionalidades individuales auto-interesadas. La ausencia de restricciones en la actuación de las partes asegura la inexistencia de pre-supuestos morales en las premisas del contractualismo liberal, y la optimalidad del resultado (es decir, la imposibilidad de justificar individualmente modificación alguna en el mismo, si se acepta el punto de partida) hace innecesaria la introducción de límites morales ulteriores.

En segundo lugar, si cabe pensar una estructura de interacción racional que logra al mismo tiempo la maximización y la optimización sin necesidad de mecanismo restrictivo alguno, entonces la moralidad aparece como una necesidad contingente, debida al infortunado hecho de que las estructuras de interacción reales no operan como el mercado ideal, y generan inevitablemente fallos que la cooperación debe corregir. Por esto puede Gauthier afirmar que "la moral surge de los fallos del mercado".

Por último, el mercado hace resaltar las características de la interacción natural. Como ya hemos señalado, el mercado es como un telón de fondo sobre

---

sociedad liberal tal como la entiende Gauthier. Pero en ese nivel, el mercado es *producto* del acuerdo moral, no su requisito o su punto de partida.

[249] Un resultado individual y colectivamente racional; es decir, un ejemplo opuesto al del Dilema del Prisionero, en el que la maximización individual prohíbe la racionalidad colectiva (optimalidad) del resultado.

el que se recorta el contorno problemático de la interacción natural, aparecen diáfanas sus causas y el papel corrector de la moralidad se discierne con mayor facilidad. El mercado ideal representa, en este sentido, el ideal de armonización entre la racionalidad individual y la colectiva. Es una guía para la cooperación, cuya misión es superar los fallos del mercado real para aproximar sus resultados a los del ideal.

Somos conscientes de que esta triple función de la idea de un mercado perfectamente competitivo puede no aparecer clara por ahora. Intentaremos ofrecer la descripción del mercado que podría clarificarla, a fin de arrojar un poco de luz sobre lo que hemos simplemente enunciado.

El mercado ideal que imaginamos se identifica casi totalmente con la situación de competencia perfecta tal como fue concebida por los economistas clásicos siguiendo a Adam Smith. Gauthier es consciente de que la competencia perfecta sólo es posible en un contexto del que estén ausentes la fuerza y el fraude[250]. Desde esta perspectiva, el mercado *no puede* considerarse una especie de "estado de naturaleza" o *base-line* previa al contrato, como equivocadamente lo han tomado algunos críticos[251]. A modo de excurso añadiremos que, según nuestro punto de vista, la situación original no puede entenderse siquiera como cierto tipo de mercado real, pese a la declaración de Gauthier de que la moral surge de los fallos del mercado. Creemos que también aquí el mercado juega un papel metafórico: los "fallos del mercado" representan la contradicción entre la racionalidad individual y la colectiva ya ejemplificada por el Dilema del Prisionero y los otros dilemas de la racionalidad. Es esa contradicción —característica de cualquier interacción entre agentes racionales auto-interesados— la que define la situación original y hace racional el tránsito hacia la cooperación.

El mercado supone, por tanto, la presencia de cierta "justicia subyacente", aunque ésta se reduzca a evitar la fuerza y el fraude en las transacciones.

---

[250] Cfr. *MA*, p. 85 y "No Need for Morality: The Case of Competitive Market", cit., p. 42.

[251] Cfr., p. ej., Danielson, P., "The Visible Hand of Morality" (en *Canadian Journal of Philosophy*, v. 18, nº 2, Junio 1988, pp. 357-384), p. 366.

Pero entonces, ¿cómo se puede afirmar que el mercado es una "zona exenta de moralidad"? La respuesta es simple: fijándonos, no tanto en las condiciones que hacen posible la interacción de mercado, sino en el funcionamiento de la interacción misma, en la *operación* del mercado. La justicia subyacente (sobre la que una teoría moral contractual tendrá mucho que decir) es tomada simplemente, para este propósito, como una condición necesaria de la estructura que llamamos mercado; pero si valoramos la estructura como tal, y su funcionamiento, encontraremos que podemos caracterizarla idealmente de modo que en ella, los agentes actúan *sin restricciones morales*[252]. En definitiva, la moral como restricción *interna* a la maximización no existe en el mercado. En el mercado perfecto no toda interacción maximizadora está permitida (pues hay límites impuestos por la estructura misma), pero toda interacción permitida puede ser maximizadora sin límites internos.

La operación que analizamos es la del mercado idealmente competitivo, el cual requiere una serie de condiciones contrafácticas que pasamos a enumerar con cierta brevedad, pues responden a la caracterización clásica de la competencia perfecta, que puede hallarse en cualquier manual de economía:

El mercado es básicamente un mecanismo para decidir la producción y distribución de bienes... y males. Los bienes del mercado son sus productos, encaminados a satisfacer la demanda individual. Los males son los factores de producción, principalmente el trabajo. Cada individuo en el mercado desea

[252] La misma idea puede verse desde esta otra óptica: los límites que deben observar los agentes en el mercado son todos externos (legales); no hay restricciones internas al auto-interés que podamos identificar con la moralidad (que caracterizan precisamente a la cooperación). Un agente moral (inmerso en una sociedad justa, esto es, una sociedad entendida como una empresa cooperativa para el beneficio mutuo basada en la moral por acuerdo) se encontrará en ocasiones obligado (moralmente) a restringir su comportamiento maximizador para sufrir un coste neto de utilidad como consecuencia de algunas de sus acciones (p. ej., ayudar a un accidentado, pagar impuestos, etc.); tal situación es desconocida para un maximizador en un mercado perfectamente competitivo: si restringe sus afanes maximizadores (p. ej., absteniéndose de defraudar o de coaccionar) ello no es a causa de que se sienta moralmente obligado, sino porque la estructura legal del mercado prohíbe estas actividades (no entramos en el circular debate de si la aceptación de límites legales supone ya un compromiso moral o no, etc.; aquí no se trata de discutir el carácter del *homo oeconomicus*, ni de caer en la misma trampa que la mayoría de los críticos, pasando imperceptiblemente del mercado ideal a nuestros análisis empíricos y nuestros prejuicios ordinarios sobre los mercados reales, se trata simplemente de valorar la operación del mercado).

tantos bienes como sea posible y tanto de ellos como sea posible (es un maximizador); pero también desea ofrecer servicios (trabajo) en la menor medida posible. La utilidad marginal decreciente de los bienes del mercado y la frontera tecnológica sitúan el nivel de producción de los bienes allí donde coste y beneficio marginal se igualan[253]. Por otro lado, la ley de oferta y demanda (en condiciones de información perfecta) permite una distribución óptima de los bienes y servicios: cualquier mejora en la situación de un individuo significa el empeoramiento de la situación de otro. Los niveles de producción y distribución del mercado se sitúan, por tanto, en equilibrio.

El intercambio económico en condiciones de libertad asegura el mantenimiento de la optimalidad y la tendencia constante al equilibrio. Para ello, hay que suponer que cada agente es un propietario individual (tanto de su capacidad de trabajo y talentos como de otros bienes que podemos llamar su "dotación inicial de factores de producción") y que *todos* los bienes son propiedad privada de alguien. En el mercado ideal *no hay bienes libres* (como son el aire o el mar en el mercado real). También el consumo ha de ser estrictamente privado: en el mercado ideal *no hay bienes públicos* que puedan ser consumidos por varios individuos al mismo tiempo. Cada individuo "paga" todo aquello que consume y consume completamente aquellos bienes por los que ha pagado. Por último —como consecuencia de la privacidad del consumo— nadie obtiene utilidad de lo que otro consume, es decir, los individuos son mutuamente desinteresados.

---

[253] La idea de marginalidad quizá requiera una breve explicación para quienes no estén familiarizados con el enfoque económico. En economía, el concepto de marginalidad se emplea para referirse a la influencia que tiene unidad o elemento adicional en el conjunto del resultado económico: así, por ejemplo, se habla del coste marginal de un producto para referirse a cuánto aumenta el coste de producción por el hecho de producir *ese bien adicional*, el beneficio marginal de un bien será, por tanto, el aumento de beneficio que se obtiene por el hecho de producir un bien más. El uso de la idea de costes y beneficios marginales es el siguiente: se supone que si un empresario tiene que decidir qué cantidad de un bien producir considerará qué beneficio marginal obtiene con cada unidad adicional, y qué coste marginal implica su producción; mientras el beneficio marginal exceda al coste marginal, producirá la unidad adicional del bien, pero cuando éstos se igualen ya no producirá más unidades adicionales (pues para la siguiente unidad, el coste marginal superaría al beneficio). El punto en que se igualan beneficio y coste marginales determinará la cantidad de bienes que se producen: producir más o menos bienes sería sub-óptimo. Este mismo análisis se aplica a otros ámbitos como la producción de bienes públicos, la distribución de beneficios cooperativos, etc.

Se puede apreciar que el mercado perfectamente competitivo es una construcción conceptual completamente contrafáctica; ajena, en todos los aspectos, a las condiciones de los mercados reales. En cuanto a los agentes que operan en el mismo, sus rasgos relevantes en cuanto participantes en el mercado son sus funciones de utilidad (que miden sus preferencias) y sus "dotaciones de factores". Estos dos datos (función de utilidad y dotación de factores) configuran lo que Gauthier denomina el "yo de mercado" de cada individuo[254]. Si un individuo se identifica con su yo de mercado[255], aceptará el resultado de la interacción de mercado como satisfactorio, pues representa lo *máximo* que puede obtener *dada* su dotación inicial y sus preferencias. Con otras palabras, si suponemos que los factores que cada individuo ofrece al intercambio son su contribución al mercado, y los bienes que obtiene son el beneficio que extrae, se puede decir que cada individuo participará en el mercado hasta que su contribución y beneficio marginales se igualen, esto es, hasta el punto en que su beneficio es máximo. La posibilidad que el mercado ideal oferece de igualar contribución y beneficio marginales garantiza un resultado óptimo tanto desde el punto de vista individual como colectivo.

Las condiciones del mercado antedichas, unidas a la información perfecta

---

[254] Pese a sus similitudes (racionalidad maximizadora, desinterés mutuo, etc.), el "yo de mercado" no puede identificarse con los agentes en estado de naturaleza (previo al contrato) porque el "yo de mercado" incluye un componente que es ilegítimo suponer pre-contractualmente: la distribución inicial de la propiedad privada. Este comentario es sugerido por la lectura de Danielson (Cfr. "The Visible Hand of Morality", cit., p. 369) cuando sostiene que Gauthier identifica las definiciones del agente necesarias para el mercado y para la cooperación. El propio Danielson reconoce más abajo que "los contratantes necesitan ciertos derechos individuales pre-contractuales, pero no necesitan derechos de propiedad completamente desarrollados" (*ibidem*). En nuestra opinión, incluso esa afirmación es dudosa: cualquier referencia a derechos como tales puede considerarse post-contractual. Más abajo defenderemos esta nuestra interpretación del contractualismo liberal de Gauthier.

[255] La identificación con la función de utilidad no presenta problemas, pues está basada en las preferencias. La identificación con la "dotación de factores" presenta, sin embargo, muchos (que no corresponden a este momento de la discusión). Se trata de los problemas de justicia que aparecen al considerar cuál debe ser la distribución inicial de factores en el mercado. Para los propósitos presentes, tomaremos las dotaciones iniciales como dadas. Más adelante se mostrará bajo qué condiciones es razonable para un individuo identificarse con su dotación inicial de factores. Insistimos una vez más que se trata únicamente de valorar la operación del mercado, es decir, la racionalidad de sus resultados (*outputs*) dada la distribución inicial de bienes y factores (*inputs*).

(especialmente concerniente a las funciones de producción que representan una tecnología dada[256]) permitirían que la producción y el intercambio se llevasen a cabo en condiciones de certeza. Cada individuo podría tomar sus decisiones sin necesidad de cálculos estratégicos sobre qué piensan hacer los demás. El mercado perfectamente competitivo es un contexto de decisión paramétrico. El principio de maximización puede aplicarse, por tanto, sin restricciones. Dado que el contexto es paramétrico, el eventual fracaso de un agente en su afán maximizador sólo puede ser achacado a él mismo. Con el símil del que gusta Gauthier, en el mercado ideal cada agente es, si se identifica con su "yo de mercado", como un Robinson Crusoe en su isla: sus deliberaciones sólo tienen en cuenta el coste y beneficio implicado por las acciones que puede realizar; si no logra realizar las más satisfactorias según su orden de preferencias (o igualar coste y beneficio marginales), él solo es culpable.

Según los economistas, la actividad libre bajo condiciones de certeza, como la descrita, hace que el mercado se mueva siempre hacia el equilibrio y, dada la competencia perfecta, ese equilibrio es un óptimo de Pareto: nadie podría consumir más productos dados los servicios que ofrece, u ofrecer menos servicios dados los bienes que consume, a no ser que otra persona accediese a producir lo mismo y consumir menos o consumir lo mismo y trabajar más.

En definitiva, las condiciones del mercado perfectamente competitivo muestran las características de un tipo de interacción en el que la persecución del beneficio individual promociona el interés de la sociedad, al producir un resultado *óptimo* mutuamente beneficioso. Como escribe Gauthier: las condiciones de la competencia perfecta hacen visible la invisible mano de que hablara Adam Smith.

Concebida la moralidad como una restricción al auto-interés en aras del interés mutuo, debe resultar evidente que en un mercado perfectamente competitivo no hay lugar para la moralidad. Gauthier concreta en tres

[256] Este punto es importante. Un mercado ideal supone una tecnología que no evoluciona: el cambio tecnológico origina externalidades de un tipo peculiar que no serían evitables por el hecho de suponer la estricta privacidad de la propiedad y el consumo. El equilibrio y optimalidad logrados por el mercado serían distorsionados por estas externalidades.

caracteres del mercado perfecto sus credenciales como zona moralmente libre: se trata de la actividad libre de los agentes, su imparcialidad respecto a los individuos y la optimalidad del resultado.

La libertad en el mercado es completa en el sentido de que nadie está en situación de imponer los términos de la interacción a otros. Tales controles surgen naturalmente en el mercado al desarrollarse monopolios o cárteles. Pero en las condiciones del mercado perfectamente competitivo tales fenómenos no ocurrirían. Cada agente en el mercado estima que la única restricción a su actividad está impuesta por sus capacidades y talentos, así como por su dotación de factores. Pero no se puede argumentar que estos límites supongan una merma de libertad. La libertad está garantizada en tanto cada agente puede poner sus capacidades al servicio de sus preferencias sin interferencia alguna[257].

La segunda credencial del mercado como zona moralmente libre es la imparcialidad que exhibe. Como nadie está en situación de aprovecharse de otros (dados los requisitos ideales, especialmente el de información perfecta), las posibles desigualdades no pueden considerarse más injustas que las desigualdades entre dos individuos totalmente aislados uno del otro. Al igual que un "robinson" en su isla disfruta de *todo* el producto de su trabajo, también un agente en el mercado perfecto puede esperar disfrutar de *todo* el beneficio equivalente a su contribución al producto total del mercado. Es decir, cada uno espera recibir tanto como da[258].

---

[257] Cfr. Gauthier, *MA*, p. 90.

[258] Este aspecto ha sido criticado (justamente, en nuestra opinión) por Daniel M. Hausman ("Are Markets Morally Free Zones?", *Philosophy and Public Affaires*, vol. 18, nº 4, otoño 1989, pp. 317-333) y Jean Hampton ("Can We Agree on Morals?", *Canadian Journal of Philosophy*, vol. 18, nº 2, junio 1988, pp. 331-356). Ambos inciden en un problema que Gauthier trata quizá con superficialidad: el problema de aquellos agentes del mercado que reciben, por sus talentos especiales (como las estrellas deportivas), mucho más del mínimo que exigirían para realizar su actividad. Estos casos distorsionan la idea de que cada uno recibe del mercado cuanto da. Gauthier evita el problema argumentando que estas personas no tienen derecho a la parte de sus rentas que excede la cantidad por la que estarían dispuestos a trabajar. Hampton (art. cit., p. 339 y ss.) y Hausman (art. cit., p. 324) contradicen solventemente el argumento de Gauthier. Según Hausman, por ejemplo, este tipo de rentas atípicas en un mercado idealmente competitivo haría que no pudiera ser considerado moralmente neutro (según el estándar de Gauthier). Sin embargo, se puede defender, con Hampton, que este tipo de rentas no tienen por qué distorsionar la imparcialidad del

Por último, la coincidencia de maximización y optimalidad que se da en el mercado (lo que le convierte en la estructura antitética del Dilema del Prisionero) elimina el fundamento de cualquier restricción acordada que nos permitiera distinguir lo correcto y lo incorrecto. Por tanto, las distinciones morales carecen de fundamento racional en el mercado perfectamente competitivo.

Como conclusión, el mercado ideal puede, según Gauthier, afirmarse como un contexto de interacción carente de moralidad y en el que no se halla fundamento para la misma:

> "Podemos concluir que la libre actividad, la ausencia de externalidades y la optimalidad son en conjunto suficientes para vencer todas las acusaciones de parcialidad en la operación del mercado —y asegurar que [...] cualquier alternativa al resultado del mercado, bien dejaría a alguien en peor situación, bien sería parcial, al permitir que algunos se beneficiaran a expensas de otros. La idea de una zona carente de moralidad queda afirmada."[259]

Ya hemos comentado que el concepto de un mercado idealmente competitivo es uno de los puntos de la obra de Gauthier que más críticas ha suscitado. Nosotros creemos que, si bien su enunciación en *MA* da lugar a confusiones y justifica muchas de las críticas recibidas, el *sentido* de esta

---

mercado. Ella ofrece, de hecho, una explicación que, por así decir, las integra con el resto de los beneficios "imparciales" del mercado. En todo caso, consideramos que se trata de una disquisición un tanto irrelevante dado el restringido y heurístico papel que, según nuestra interpretación, se debe asignar al mercado. Para nuestros efectos, la explicación de Hampton puede (en este punto) solventar los problemas de la argumentación de Gauthier y permitirnos mantener nuestra interpretación en lo demás.

[259] Gauthier, *MA*, pp. 98-99. Esta misma conclusión es expresada con otras palabras en el artículo de 1982: "Los resultados del mercado no son moralmente correctos ni incorrectos. Proponemos defender que las características de nuestra concepción común de las personas como agentes racionales individuales [...] y de nuestra concepción común de la moralidad como una restricción imparcial a la persecución individual del auto-interés o beneficio, conducen a la conclusión de que la valoración moral está restringida a la actividad no-de-mercado" ("No Need for Morality: The Case of Competitive Market", cit., p. 47).

construcción ideal es muy interesante. En nuestra opinión, Gauthier quiere simplemente emplear la idea económica de competencia perfecta (con resultado equilibrado y óptimo) como antítesis de la estructura de interacción ejemplificada por el Dilema del Prisionero. La interacción ideal de mercado es una construcción heurística como lo es el estado de naturaleza. Lo relevante es que, siendo elementos constructivos (individuos y racionalidad) los mismos que los de éste —la única diferencia estriba en que se suprimen ciertas condiciones fácticas, como los costes de información, las externalidades, la no-escasez de ciertos bienes, etc.— sin embargo, el resultado es justamente opuesto. La conclusión que, a nuestro juicio, Gauthier quiere extraer de este ejercicio podría enunciarse así: postulados los elementos del estado de naturaleza, el conflicto que hace necesario el pacto moral surge, en efecto, *pero pueden pensarse las condiciones bajo las cuales no habría surgido*, sin renunciar a esos postulados.

De acuerdo con esta conclusión, las causas del conflicto en el estado de naturaleza se sitúan, no tanto en el carácter egoísta de los individuos (como pensó Hobbes), sino en el hecho de que se den en el mundo (representado por el mercado real) ciertas condiciones fácticas que desarmonizan la racionalidad individual y colectiva.

Porque, en efecto, quizá lo que más resalta conforme se van enumerando las condiciones necesarias para la competencia perfecta, es que no tiene nada que ver con el mundo real. La diferencia no es sólo "técnica": no es que en el mundo real sea imposible contar con información perfecta o calcular el coste exacto de usar un bien público; la diferencia es ontológica. La competencia perfecta requiere supuestos rigurosamente contrafácticos: no se dan, ni pueden darse, en el mundo tal como lo conocemos.

Por eso, el aparente panegírico de la neutralidad del mercado que se encuentra en los textos de Gauthier debe entenderse limitado a la función heurística que hemos visto que juega en una especie de articulación o coyuntura de su teoría. No se trata en absoluto de una defensa de la economía ultra-liberal; porque, como escribe Gauthier, "nos interesa mostrar que la moralidad no tiene lugar en un contexto ideal de interacción, no sostener que este ideal

tenga una aplicación práctica directa"[260]. Así, quienes, como Hausman, montan su crítica sobre la base de que se puede parangonar el mercado perfecto con las condiciones de mercado habituales (o simplemente se deslizan hacia tal identificación) malinterpretan completamente el alcance del concepto de mercado en la teoría de Gauthier.

Una segunda consecuencia del carácter contrafáctico del mercado ideal es que, si bien éste se muestra como "solución" a la contradicción entre la racionalidad individual y colectiva, el mercado real exhibe esta contradicción con toda su crudeza[261]: los "fallos del mercado" no son sino expresión de esta insoluble contradicción, ejemplificada en los problemas de las externalidades[262], del *free-rider*[263], el dilema del contribuyente[264], etc.

---

[260] "No Need for Morality...", cit., p. 48.

[261] Gauthier reconoce la distancia entre el mercado ideal y el real en un texto especialmente claro: "Pero por muy iluminador que sea el mercado para mostrarnos la posibilidad de interacciones que no dan lugar a problemas para maximizadores de valor relativo al agente —de hecho, el mercado es iluminador porque nos revela un tipo de interacción que no necesitaría ser guiada por esos principios restrictivos del comportamiento maximizador que constituirían una moral racional— sin embargo, a muchos de nosotros el mundo real no nos parece una aproximación muy cercana al reino de la competencia perfecta" ("The Incompleat Egoist", en Gauthier, D., *Moral Dealing*, Ithaca, Cornell U.P., 1990, pp. 234-273; p. 260).

[262] Se denominan externalidades, "costes desplazados" o "economías externas" a aquellos efectos económicos que no repercuten directamente en aquél que realiza la actividad que los produce. El ejemplo clásico es la polución: el empresario que vierte sus desechos a un río ocasiona un coste a la sociedad; desplaza parte del coste de producción de sus bienes a los usuarios del río, sin asumirlo él mismo. A consecuencia de esta externalidad, la oferta de estos bienes contaminantes es superior a la que se daría en un mercado perfecto, en el que el productor (y, por tanto, cada consumidor de ese bien) tuviera que asumir todo el coste de producción. Aunque tienen menor importancia económica, también pueden darse externalidades positivas, cuando la actividad de un agente beneficia a otro u otros sin proponerselo. El ejemplo característico es la construcción de un faro cuya luz, una vez construido, beneficia a cualquiera que por allí pase,haya contribuido o no a su producción. Al igual que los bienes contaminantes tienen a estar sobre-ofertados, los bienes que generan externalidades positivas tienden a ser infra-producidos por el mercado.

[263] *Free-rider*, gorrón o parásito se denomina a aquél individuo que pretende beneficiarse de los bienes públicos sin contribuir a su producción. La conducta "gorrona" es la individualmente racional en un mercado que, de hecho, proporciona bienes públicos. Con el ejemplo del faro, es individualmente racional no contribuir a su construcción, pues ello no impide beneficiarse, después, de su funcionamiento.

El análisis paralelo del mercado ideal y el real muestra, por tanto, que, eliminadas las imposibles condiciones para la competencia perfecta, los mercados fallan en muchos casos. Hampton, quien ha formulado esto con especial claridad, escribe que "el mercado fracasa siempre que las actividades racionales de intercambio entre personas racionales no llevan a las partes a un resultado óptimo de Pareto"[265]. Muchas interacciones en el mercado real tienen la estructura de un Dilema del Prisionero, en el que "cada agente tiene una estrategia estrictamente dominante, y si todos la usan, lo que consiguen es un estado estrictamente inferior en términos paretianos. Son conducidos, de un modo que habría sorprendido a Adam Smith, por una malévola mano invisible a promover un fin que no formaba parte de su intención y que ninguno de ellos deseaba"[266]. Así ocurre en los dilemas que hemos mencionado arriba, que básicamente se reducen a la producción de bienes públicos: todos prefieren que haya bienes públicos a que no los haya, pero, si cada uno obedece su auto-interés, los bienes públicos no se producirán. Una infinita cascada de estudios sobre la elección colectiva ha demostrado que se trata de verdaderos dilemas: excepto con medios coactivos, no hay modo de armonizar los intereses individuales y los colectivos, incluso aunque se reconozca el beneficio individual de la producción de bienes públicos.

Pues bien, allí donde la interacción de mercado conduce a dilemas que hacen fracasar uno de los objetivos de la racionalidad, como es la optimización, ha de entrar en juego un nuevo modo de interacción racional: la interacción cooperativa.

---

[264] El dilema del contribuyente viene a expresa un caso de conducta parásita. El razonamiento de un contribuyente egoísta sería el siguiente: si todos los demás (o un número suficiente) contribuyen, el bien público se producirá, así que haré mejor en no contribuir pues me beneficiaré del bien público de todos modos; si nadie (un número insuficiente) contribuye, el bien público no se producirá, por lo que no habrá beneficio, con lo que contribuir sería irracional (sería contribuir "para nada"). En conclusión, hagan lo que hagan los demás, la conducta individualmente racional es no contribuir.

[265] Hampton, J., "Can We Agree on Morals?", cit., p. 334.

[266] Watkins, J., "Second Thoughts on Self-Interest and Morality" (en Campbell, R y Sowden, L. (eds.), *Paradoxes of Rationality and Cooperation: Prisoner's Dilemma and Newcomb's Problem*, University of British Columbia Press, 1985, pp. 59-74), p. 59.

Por eso se afirma que la moral surge del fracaso del mercado. Con ello, se quiere indicar que allí donde la estructura de la interacción natural entre individuos maximizadores auto-interesados conduce a un resultado sub-óptimo, está justificado establecer mecanismos de restricción de la conducta maximizadora, si con ello se consigue un resultado óptimo, es decir, más beneficioso, *individualmente*, para todos.

Así es como, sin transformar los postulados iniciales de la teoría, se puede pasar del esquema ideal de interacción competitiva, al esquema de interacción real y, desde él, justificar la necesidad de restricciones que previamente hemos identificado con la moralidad. Este razonamiento muestra que, a pesar de que es pensable una zona exenta de moralidad, las condiciones de la interacción (cuyo análisis es facilitado por la construcción hipotética de ese contexto libre de moralidad) conducen naturalmente —siguiendo los dictados de la racionalidad individual maximizadora— a demandar la cooperación para superar sus dilemas.

Como decíamos al inicio de este punto, nos parece que la idea de un mercado perfecto y su puesta en relación con la interacción real para justificar de modo intuitivo la necesidad de la cooperación racional, si bien no es parte esencial del argumento contractualista, ayuda a percibir el papel de la moralidad. La moralidad nace instrumentalmente (de hecho, no nace como tal moralidad, sino como simple cooperación) para solucionar los problemas de la racionalidad estratégica. Y viene requerida, no por compromisos normativos previos, ni por sentimientos empáticos, ni por una subjetividad moral anterior, sino por el mismo afán maximizador racional que provoca los fallos en el mercado.

Tal como lo interpretamos, el mercado es sólo una imagen o un símbolo, pero una imagen muy pertinente, que despeja dudas en el camino hacia la deducción de una moralidad racional.

Desgraciadamente, la exposición de la idea de un mercado perfectamente competitivo en *MA* da lugar a confusiones que, como ya hemos apuntado, suscitan diversas críticas. Tal vez Gauthier no hace suficiente hincapié en la distinción entre el papel heurístico y referencial de una zona exenta de

moralidad, el papel paradigmático o explicativo de la alusión a los "fallos del mercado" real, y el papel teórico del estado de naturaleza, representado en su teoría por la "posición inicial de negociación". En estos tres ámbitos los protagonistas son los mismos: agentes racionales auto-interesados. Pero la situación en que les colocamos difiere notablemente: en la situación de competencia ideal se les sitúa en un marco de interacción contrafáctico diseñado para nuestros propósitos heurísticos. La valoración moral de este marco está fuera de lugar. Únicamente interesa valorar la operación y los resultados dadas sus condiciones iniciales. En la situación real de mercado sí cabría valorar aspectos tales como la distribución inicial o la institución misma, y precisamente esa es una de las funciones de la teoría moral: establecer una norma para valorar la justicia de instituciones como el mercado, la propiedad privada, etc. Sin embargo, en esta fase de la teoría, la mención del mercado real tiene otro alcance, pues sirve simplemente para mostrar que, situados los agentes en un marco de interacción real, las condiciones que permitían hablar de un reino carente de moralidad (en el mercado ideal) desaparecen, y la racionalidad individual misma demanda un modo de interacción capaz de superar el fracaso que supone la divergencia entre maximización y optimización. Por último, la posición inicial de negociación no se identifica con el mercado real porque en éste se supone que los agentes poseen propiedades y comparten una estructura institucional de derechos (que al hablar del mercado ideal llamábamos "justicia subyacente"), mientras que la posición inicial es *anterior* a toda institución cooperativa y a todo esquema de derechos[267]. Por lo tanto, en la posición

[267] Desde el punto de vista de la teoría contractual, el mercado es una *institución cooperativa*, que es tanto como decir que es una de las instituciones sociales diseñadas para el beneficio mutuo que puede surgir del pacto fundante. Esto contrasta con la contraposición entre interacción de mercado e interacción cooperativa, que venimos empleando (en la que se basa la afirmación de que cuando el mercado falla es necesario dejar paso a la cooperación). Esta contraposición es posible porque, desde el comienzo de la discusión sobre el mercado, hemos dejado aparte la consideración del mercado globalmente, como institución, para centrarnos en su operación o funcionamiento interno, y ese funcionamiento no es cooperativo, sino competitivo; pero eso no excluye que el mercado, tomado globalmente, sea una institución cooperativa (sobre este punto, cfr. Braybrooke, D., "Social Contract Theory's Fanciest Flight", cit., p. 757). Este doble punto de vista debería aclarar aún más la diferencia entre el mercado (sea ideal o real) y el estado de naturaleza: el mercado no sirve como ejemplo del estado de naturaleza porque es una institución netamente post-contractual. Otra cosa es que el modelo de interacción competitiva de mercado sirva para definir los contornos de la interacción cooperativa.

inicial no se supone ni propiedad, ni sistema de intercambio, ni ninguna de las restantes características que conforman el mercado. El marco de interacción en que situaremos a los agentes es mínimo: sus elementos se reducen a la racionalidad individual perfecta unida a la información completa.

La neta distinción de estos tres ámbitos habría evitado probablemente críticas como las de Peter Danielson, quien identifica el mercado con la situación pre-contractual[268]; o Jean Hampton, que trata la teoría de la negociación de Gauthier como un sistema de distribución para los casos en que el mercado no la realiza satisfactoriamente.

En general, los primeros comentaristas de la obra de Gauthier parecen, en lo que concierne al mercado, o bien fascinados por la idea de tratar la moral como un bien público que el mercado naturalmente no produce pero que es colectivamente beneficioso producir; o bien anclados en una lectura económica, según la cual la teoría de Gauthier pretende sencillamente solucionar el sempiterno problema de la producción de bienes públicos.

En cualquiera de los dos casos aplican los conceptos analíticos usuales en el tratamiento económico de los problemas suscitados por la producción de

---

Por abundar en este punto, pondré un ejemplo: imaginemos un grupo de clubes deportivos que deciden asociarse para disputar una liga. Lógicamente el funcionamiento de la liga es competitivo (de otro modo dejaría de cumplir el objetivo para el que se crea), pero a esa actividad competitiva debe preceder una actividad cooperativa: la organización de la misma. Ello no impide que la disputa deportiva post-contractual, pueda tomarse como modelo más o menos aproximado del tipo de relación que había entre los clubes antes de asociarse para organizar una liga. Tan absurdo es creer que la relación de los clubes antes de fundar la liga era deportiva como creer que la interacción natural entre agentes auto-interesados es una interacción de mercado.

Debemos añadir, por último, que somos conscientes de que las ideas expuestas en esta nota (la concepción del mercado como un "bien público" en sí mismo) no aparecen, o lo hacen de modo oscuro, en *MA*. De todas formas, se trata de una interpretación derivada del sentido general de la teoría contractual de Gauthier, y acorde con sus últimos comentarios sobre la misma.

[268] Cfr. Danielson, P., "The Visible Hand of Morality", cit., pp. 366 y ss. Según Danielson, el argumento de Gauthier no seguiría los pasos que exponíamos arriba (en los que el mercado no tiene un papel crucial), sino estos otros: 1) se identifica el estado de naturaleza con el mercado; 2) se descubre que, si el mercado fuese idealmente competitivo, produciría la cantidad óptima de todos los bienes y la moral no tendría razón de ser, pero, como el mercado tiene fallos, hay bienes que no se producen: los bienes públicos; 3) el pacto permite asegurar la producción de bienes públicos (y versa sobre ellos únicamente, dejando la producción de bienes privados al mercado) e introduce restricciones morales.

bienes públicos. Se trata, sin duda, de un enfoque posible, pero extraordinariamente parcial. Con nuestra interpretación "moderada" del papel del mercado en la teoría de Gauthier pretendemos acercarnos más al sentido de la obra, que es el de reconstruir racionalmente los principios de la moral. El empleo en esta reconstrucción de materiales provenientes de la economía y la teoría de juegos (que se muestran por lo demás muy válidos y potentes para el análisis), no nos autoriza a sesgar tanto la lectura que ésta quede reducida a un debate sobre la producción de bienes públicos. Como ha señalado David Braybrooke, tal vez el mejor enfoque consiste precisamente en ver el mercado mismo como un bien público para, de este modo, resaltar el papel fundamental del argumento contractualista que, en los textos de otros críticos, es trivialmente interpretado como un expediente técnico para intentar superar los fallos del mercado mediante la producción y distribución de bienes públicos.

El origen de la generalizada trivialización del papel del mercado en la teoría de Gauthier puede tener su origen en el texto mismo de *MA*: primero, por su realismo al describir económicamente la operación del mercado, que ha conducido a la mayoría de los críticos a acabar enredados en una discusión de los mercados reales, sin percibir el carácter heurístico o "metafórico" que el concepto tiene en *MA*. En segundo lugar, por que Gauthier no distingue suficientemente los ámbitos respectivos de la cooperación y del mercado. En ocasiones parece afirmar que la producción de bienes privados debe quedar en manos del mercado (como si éste fuera una institución "natural" que, como produce y distribuye los bienes privados óptimamente, no necesita justificación ni modificación alguna) mientras la producción y distribución de bienes públicos es el ámbito de la cooperación y el acuerdo (el ámbito de la moralidad, por tanto). Siguiendo con este argumento, nos encontraríamos que el contrato social no afecta a aquella parte de la interacción en que la competencia es el mejor modo de decidir sobre la producción y distribución de bienes y servicios, sino sólo a la parte que tiene que ver con la producción de bienes públicos. Según esto, se podría afirmar, con un contenido más substantivo, que el mercado es una zona (naturalmente) carente de moralidad, y así debe dejarse, como una especie de "parque natural protegido" en medio de la cooperación nacida del contrato. La sociedad liberal post-contractual quedaría, así, dividida

en dos partes: el mundo de la interacción cooperativa, y el mundo de la competencia, que sería un resto de "estado de naturaleza". Tal concepción de la sociedad y del papel del contrato representaría la cara más liberal del pensamiento de Gauthier. Una cara que, a pesar de dejarse traslucir en algunos textos, no responde al significado de su obra, que enfatiza el componente cooperativo (el propiamente contractual) de la sociedad como un todo.

Porque, en efecto, en los textos más explícitos, aparece claramente el papel puramente heurístico del mercado como situación de interacción ideal. De hecho, en varios momentos, Gauthier niega expresamente que él defienda una política económica liberal. El mercado es calificado como una institución compleja que, para existir, debe ser elegida por los agentes en la situación original[269]. Claramente afirma Gauthier que el mundo no es un mercado y, precisamente por ello, las restricciones morales son necesarias. Desde este punto de vista, no habría un doble estándar de interacción en una sociedad liberal. Una sociedad liberal, justificada contractualmente, es concebida, en su conjunto, como una empresa cooperativa para el beneficio mutuo. Tal vez incluya entre sus instituciones la compleja estructura del mercado (que implica la creación de un ámbito de libre competencia), porque los individuos entiendan que es el mejor sistema para llevar a cabo sus decisiones sobre la producción y distribución de bienes privados. Pero en todo caso ése es un asunto de decisión colectiva: tal vez no haya lugar para el mercado en la sociedad post-contractual.

En conclusión, nuestro análisis muestra que el papel del mercado no prejuzga la opinión sobre cómo ha de ser una sociedad justa, ni supone necesariamente defender una estructura económica concreta. De un modo que no han alcanzado a ver la mayor parte de los críticos, el mercado se sitúa, en el argumento contractualista moral, como un fondo oscuro sobre el que resaltan, luminosos, una serie de conceptos nucleares: la perfecta racionalidad de las partes, la ausencia de pre-concepciones morales en su caracterización, la naturaleza ontológicamente contingente de la moralidad, su relación con la

[269] Cfr. p. ej., *MA*, p. 84.

producción de bienes públicos, la estructura dilemática y contradictoria de la interacción natural, etc. De este modo el mercado, sin ser una de las premisas del argumento contractualista, contribuye a clarificarlas todas. Es un error considerarlo como parte de la explicación de la interacción "natural", es decir, parte de la segunda fase de un argumento contractualista.

5. Premisas teóricas y estado de naturaleza

Proponíamos al inicio del capítulo ofrecer un análisis de los presupuestos filosóficos y metodológicos del contractualismo en general y del contractualismo moral liberal de David Gauthier en particular. En este análisis hemos ido, creo, concentrando progresivamente nuestra atención en las premisas teóricas que definen la primera fase del argumento expuesto en *MA*. A pesar de que es difícil aplicar la noción clásica de "estado de naturaleza" a las teorías neocontractualistas, se puede decir que los conceptos que hemos estudiado en este capítulo conforman el equivalente a lo que en la tradición contractualista suponía la definición de aquél escenario hipotético[270]. Es característico del argumento contractualista que su primera fase consista en la *estipulación* de una serie de elementos y condiciones cuya concepción varía bastante de un modelo teórico a otro. En estas páginas hemos señalado cuáles de estos elementos son compartidos por la mayoría de las teorías del contrato y qué distingue al enfoque liberal concreto defendido por Gauthier. Los argumentos y explicaciones ofrecidos muestran, según creemos, una concepción de las premisas del argumento contractualista no sólo claramente distinta y original, sino también considerablemente plausible. Nos gustaría ahora, como resumen y final del capítulo, recopilar las conclusiones más importantes que ha arrojado nuestro análisis de las premisas teóricas del contractualismo moral liberal.

Hemos mostrado, en primer lugar, que las premisas del contractualismo moral responden a un compromiso de economía teórica. La descripción del estado de naturaleza que emplea el contractualismo moral evita complejidades innecesarias y supuestos difíciles de justificar. Los rasgos conceptuales en los

[270] Debemos la formulación precisa de esta idea al perspicuo estudio de Jody S. Kraus, *The Limits of Hobbesian Contractarianism*, Nueva York, Cambridge University Press, 1993; en esp. cap. I.

que insiste son, por supuesto, postulados e ideales; pero apoyándose siempre en descripciones y explicaciones "estándar" de las ciencias sociales, e intuitivamente adecuadas. De este modo, las premisas, pese a ser hipotéticas, resultan plausibles desde el punto de vista del individuo real ante quien debe desplegarse la justificación contractualista de la moral. Se puede decir que, sin tratarse de una estipulación "realista", del tipo de las de Locke o incluso Hobbes[271], los elementos del contrato liberal reflejan suficientemente cómo somos las personas reales en ciertos aspectos y comportamientos relevantes. Y, sobre todo, se alejan del idealismo extremo representado por el contractualismo rawlsiano[272], que implica la defensa de un concepto hipotético de agente "tras el velo de ignorancia" tan abstracto, que finalmente es imposible su identificación post-contractual con los individuos reales a quienes se dirige el argumento —y sin esa identificación, el contractualismo moral pierde cualquier eficacia motivadora.

En concreto, hemos considerado la idea de las partes del contrato (individuos perfectamente racionales y auto-interesados) desde una ficticia división analítica que nos ha permitido comprobar que, en efecto, el individualismo es un postulado de la teoría (reflejo de un postulado metodológico liberal común), pero no así la racionalidad. Aunque los requisitos de racionalidad

---

[271] Ambos autores se apoyan frecuentemente en ejemplos históricos y presentan el "estado de naturaleza" como un estado pre-político tal como sería dados sus respectivos concepciones (que estimaban ciertas) de la naturaleza humana. Sobre la distinción entre "realismo" e "idealismo" en la construcción del estado de naturaleza, ver Kraus, J., *op. cit.*, p. 22. La misma distinción, formulada de otra forma ("plausibilidad fáctica" vs. "aceptabilidad normativa"), se encuentra en Reinhard Zintl, "Contrato sin presupuestos" (en Kern, L. y Müller, H.P. (eds.), *La justicia: ¿discurso o mercado?*, Barcelona, Gedisa, 1992, pp. 181-207) y en el libro de Barry, *Theories of Justice*, Londres, Harvester, 1989.

[272] Esta distanciación del idealismo rawlsiano es explicada por Barry en términos de comprensión de ambas teorías de la justicia: mientras Rawls presenta una teoría de la justicia como imparcialidad, Gauthier ofrece una teoría de la justicia como beneficio mutuo. Este distinto enfoque conlleva diferentes estructuras y bases teóricas (Cfr. Barry, B., *Theories of justice*, Londres, Harvester, 1989, parte III, caps. 8 y 9). Respecto a la caracterización de las partes, esta diferencia supone la aceptación del egoísmo "natural" por parte de las teorías del beneficio mutuo, mientras implica una mayor elaboración de una posición original por parte del enfoque de la justicia como imparcialidad. Esto explicaría que, pese a basarse ambas en premisas hipotéticas, las de las teorías del primer tipo resultan más "realistas".

perfecta, información completa, desinterés mutuo y otros, son todos ellos hipotéticos, el concepto mismo de racionalidad como maximización de la utilidad se asienta en las descripciones empíricas proporcionadas por las ciencias sociales, en particular por la Economía y las Teorías de la Decisión y Juegos. A su vez, el subjetivismo/relativismo axiológico que completa el dibujo de la situación inicial, se apoya en la teoría de la racionalidad. No obstante, el contractualismo liberal no se compromete con la carga normativa que esta teoría pueda tener en su aplicación a la economía, la política o la psicología. Acepta —y esto sólo en parte— su contenido descriptivo, pero su contenido normativo habrá de ser cuestionado por las conclusiones del análisis contractualista.

También hemos expuesto el papel heurístico del mercado en la teoría de Gauthier. Ha quedado claro que no se trata de un supuesto del contractualismo, sino de un ideal de interacción racional que puede servir para confirmar la neutralidad de las premisas y para sugerir la naturaleza de una moralidad contractual.

De acuerdo con nuestra interpretación de esta parte inicial de la teoría de Gauthier, se puede decir que los elementos básicos del "estado de naturaleza" son el individualismo, la racionalidad como maximización y el subjetivismo axiológico. Bajo nuestro punto de vista, las críticas que se han dirigido a estos elementos merecen distintas valoraciones: los argumentos contra el individualismo parten, en su mayoría, del olvido del carácter metodológico de este supuesto; la racionalidad económica puede rechazarse, desde luego, como paradigma de la razón humana —lo cual esta perfectamente justificado incluso desde el punto de vista contractualista, como veremos—, pero no puede negarse que constituye un mínimo común de racionalidad compartido por todos los seres humanos (cualquiera que sea su educación y carácter moral). Admitiendo esto, se ha criticado la versión concreta que ofrece Gauthier, pero nosotros entendemos que su misma simplicidad la defiende suficientemente de estos ataques. El aspecto más criticado ha sido, tal vez, la concepción del mercado. Sobre esto, nuestra conclusión ha sido que, en efecto, la formulación de

Gauthier no es del todo clara, pero que la comprensión del *sentido* de su uso del concepto podría haber ahorrado muchas críticas; la mayoría de ellas asentadas en confusiones injustificables.

Con toda seguridad, nuestros argumentos, interpretaciones, explicaciones y reformulaciones no persuadirán —o lo harán difícilmente— a quienes estén convencidos de la verdad de otras visiones de la razón, del valor o de la sociedad. Ante esto, quizá deberíamos unirnos al lacónico mensaje de Gauthier en un reciente artículo, donde expresa que sus investigaciones tienen valor dentro del marco de un "desacuerdo razonable", pero carecen de interés para quien crea que el desacuerdo práctico, o los conflictos entre valores subjetivos, muestran alguna imperfección o error en una de las partes enfrentadas o en ambas[273]. No se puede negar la agudeza de ese mensaje, al tomar conciencia expresamente de una oposición teórica tal vez insuperable. Sin embargo, queremos expresar nuestra convicción de que la fortaleza del pensamiento liberal reside en su flexibilidad, y que no debe abandonar el esfuerzo por hacer comprender su alcance incluso ante aquellos que lo considerarán un discurso maquiavélicamente ideológico o ingenuamente equivocado. En este sentido, hay que insistir en que la presentación de las premisas teóricas no tiene el alcance de una discusión ontológica. La aceptación, como punto de partida, del individualismo o la racionalidad instrumental no prejuzga —como dejamos especialmente claro al hablar del individualismo— la estructura de la sociedad que un liberal aceptaría como justa. Desde este punto de vista, el contractualista liberal se encuentra legitimado para solicitar, incluso a sus adversarios teóricos más contumaces, el asentimiento a unas débiles premisas hipotéticas cuyo desarrollo podría —aunque, ciertamente, es improbable— justificar sobre una base liberal precisamente aquello que esos adversarios defienden.

Precisamente la defensa más convincente que puede ofrecerse de las premisas que constituyen la primera fase del argumento contractualista liberal

[273] Nos referimos a "Assure and Threaten", *Ethics*, 104 (Julio, 1994), pp. 690-721; p. 691 (nota 2).

es su *debilidad*. Con esto se quiere decir que se trata de concepciones capaces de atraer un amplio grado de consenso porque no incluyen ni pretendidas verdades metafísicas ni prejuicios sobre la sociedad o la razón. El contractualismo extrae su fuerza de esa economía teórica pues, como escribe Zintl, "cuantas menos decisiones previas del teórico caractericen de antemano la celebración del contrato y las propiedades de los sujetos contratantes, cuantos menos prejuicios introduzca, tanto más convincente será su argumento"[274]. Gauthier parece ser quien más lejos ha llevado esta economía, al menos a juicio de Danielson, quien, refiriéndose a *MA*, opina que "lo que atrae nuestra atención son sus premisas y método radicalmente minimalistas"[275].

Este "radical minimalismo" tiene un objeto primordial: convencer al escéptico. El punto de vista contractualista es que la razón de ser de la moralidad no se cuestiona desde los sentimientos de amor al prójimo, o de cariño fraternal y familiar. La moralidad se cuestiona desde el egoísmo, desde la duda sobre sus fundamentos sentimentales, desde la crítica a la objetividad de la razón. Y es hacia esos lugares hacia donde el argumento contractualista debe dirigirse. El fundamento de este planteamiento es que, si logramos persuadir al egoísta de que ser moral le beneficia personalmente, con más motivo quedará asegurado en sus creencias el altruista. El contractualista es, por tanto, un argumento contra el escéptico (o el egoísta). Se concede al escéptico la ventaja inicial, y se intenta argumentar desde sus premisas, de modo que la conclusión sea irrefutable, incluso desde su punto de vista[276].

---

[274] Zintl, R., "Contrato sin presupuestos", cit., p. 181.

[275] Danielson, P., "The Visible Hand of Morality", cit., p. 362.

[276] En este sentido, dice Gauthier: "Puedo afirmar que mi ataque al egoísta requiere supuestos mucho más débiles que los que eran necesarios para los ataques llevados a cabo por los moralistas tradicionales. Donde ellos asaltaban sus posiciones desde fuera, intentando abatir sus premisas, yo las socavo desde dentro, mostrando que sus premisas no lo sustentan en absoluto. No necesito ni valor absoluto ni racionalidad universalizada. Puedo afirmar que los teóricos morales han recurrido a estas líneas de ataque porque no han visto la posibilidad de defender la moralidad combatiendo al egoísta en su propio terreno. Y así, puedo afirmar que el éxito de las tradiciones platónica y kantiana ha sido debido al no reconocimiento de un tercer modo en que el moralista podía recuperar el valor y la razón de las garras del egoísta." ("The Incompleat Egoist", cit., p. 269). Afirmaciones tan tajantes como las expuestas (y los argumentos que las apoyan) han conducido a

Bajo esta luz se comprende fácilmente la preocupación por no introducir supuestos morales injustificados en la concepción de las premisas. Tales supuestos prohibirían el éxito de un proyecto contractualista anti-escéptico como el que se trata de llevar a cabo[277].

Esto no quiere decir que si se demostrase que alguna de las premisas del contractualismo liberal posee un contenido o fundamento normativo, el proyecto como un todo hubiera de ser rechazado. Tal constatación no situaría a Gauthier en desventaja respecto a otros neo-contractualistas[278]. Su radical minimalismo le seguiría situando mucho más cerca del objetivo de neutralidad que todos ellos persiguen.

No obstante, un argumento como el anterior no es necesario a la luz de nuestro análisis. Hemos comprobado que, tal como afirma Gauthier, las premisas del contractualismo liberal no incluyen, hasta ahora, supuestos morales espurios. Incluso los requisitos más problemáticos, como la "igual racionalidad" admiten una interpretación en términos de maximización y auto-interés. Nada, en lo que llevamos visto, autoriza ha afirmar que se han

---

Danielson a reconocer que "la *Moral por acuerdo* de David Gauthier toma en serio la exigencia de que una teoría de la moralidad sea capaz de responder al desafío del escéptico racional" ("The Visible Hand of Morality", cit., p. 358).

[277] Este punto ha sido nítidamente señalado por Harsanyi cuando, en su crítica a Rawls, afirma que "en un momento u otro implica algunas elecciones morales altamente irracionales, que representan grandes desviaciones de la persecución común de los intereses humanos y humanistas, lo cuales constituyen, bajo mi punto de vista, la verdadera esencia de la moralidad" (Cfr. "Morality and the Theory of Rational Behaviour", cit., p. 41).

[278] Hampton sostiene la opinión de que, establecido el debate entre teorías contractualistas basadas en la imparcialidad (Rawls) y otras basadas en el mérito (Nozick), ninguna teoría contractualista puede escapar a la necesidad de ofrecer razones morales (previas) para "seleccionar un favorito" (Cfr. "Can we Agree on Morals?", cit., p. 343 y ss.), a pesar de la ingenua esperanza (generada por el propio Rawls) de que es posible justificar la moralidad sobre premisas estrictamente no-morales. Conjeturamos que, según esta opinión, ya que todo contractualismo está condenado a cierta decisión moral previa, será más plausible aquél que reduzca su relevancia al mínimo. Sin embargo, no compartimos la opinión de Hampton: la opción por uno u otro tipo de teoría de la justicia puede venir sugerida por la adopción de premisas razonables y la deducción subsiguiente. No vemos la necesidad de suponer que este proceso esté vinculado a creencias morales previas.

excedido los límites de ciertos postulados metodológicos y una concepción de la racionalidad expresamente no-moral. De hecho, la fidelidad al compromiso de deducción de la moralidad desde premisas estrictamente ajenas a la misma se mantendrá a lo largo de toda la segunda fase del argumento: la explicitación del carácter de la interacción natural y su solución a través de la negociación y el pacto.

Por último, se nos permitirá una breve observación sobre la construcción de las premisas. El hecho de referirnos a ellas como "estipulaciones plausibles" o "ampliamente aceptables" habrá producido rechazo en quienes, observando introspectivamente su carácter o sus inclinaciones personales, hallen que no tienen absolutamente nada en común con estos individuos descastados, desarraigados, egoístas, a-morales y competitivos que hemos descrito. No se trata ahora de la desencaminada objeción de que los postulados contractualistas no captan la esencia de los hombres, que ya ha sido contestada defendiendo su carácter exclusivamente hipotético. Se trata de la objeción más profunda que acepta que estos personajes hipotéticos tal vez sirvan para desarrollar la historia ficticia de un contrato ideal, e incluso para demostrar que es racional cumplir con las demandas de ese contrato, pero niega que todo eso signifique algo para nosotros, seres humanos concretos, ajenos a esa historia hipotética, sujetos cuyo carácter moral tiene fuentes bien distintas e irreductibles a un análisis "económico".

A esta objeción, conectada, como decimos, con un rechazo casi visceral de las premisas argumentales que hemos presentado, contestamos lo siguiente:

En primer lugar, el contractualismo no niega que la moral tiene su origen en la autonomía del sujeto. El contractualismo no es convencionalismo (como explicaremos en el capítulo siguiente). Nuestra teoría no discute, hasta el momento, la fuente últimamente histórica y social de la moralidad individual (y de la subjetividad misma). Tampoco discute los hechos psicológicos sobre el desarrollo del conocimiento moral o sobre la experiencia subjetiva de la moral (sentimientos morales). Es perfectamente lógico que, situados como estamos en distintos contextos y tradiciones, algunos de nosotros rechacemos instintivamente identificarnos con un "agente" con quien no admitimos tener

nada en común. Esto mismo le pasa al individuo liberal del que habla Gauthier al final de su teoría. Allí despliega, frente al descarnado agente económico, su concepción de un sujeto liberal que se reconoce libre, pero acepta, en uso de su libertad, límites racionales (reglas morales) no coactivos. Este individuo —al igual que el posible lector perplejo— no es idéntico a los agentes auto-interesados que hemos descrito. Pero —y aquí está la clave de la plausibilidad de las premisas del contractualismo— reconoce que es *razonable* identificarse con ellos, al menos como punto de partida heurístico para asegurar la justificación de la moralidad sobre las bases más firmes que sea posible. Cuanto más abarcantes, más sólidas serán estas bases. Si el concepto de racionalidad instrumental, por ejemplo, es tan simple que es aplicable incluso a otros mamíferos[279] (y se demuestra que las restricciones morales pueden deducirse de él), miel sobre hojuelas. Las concepciones más complejas (que suelen tener de su parte la autoridad de prestigiosas tradiciones filosóficas) están más expuestas (aún) a la crítica de ser el producto cultural de un determinado lugar y tiempo, con lo que su plausibilidad como base de una moralidad universal decrece. Por otro lado, el individuo liberal reconocerá —si es capaz de hacer una reflexión de buena fe, y suponiendo que puede imaginarse a sí mismo despojado de sus creencias morales— que, en aspectos fundamentales de su acción (en ámbitos enteros de su vida), guarda una estrecha semejanza con las "partes del contrato" tal como han sido definidas.

Hobbes, que temió una objeción similar a la que nosotros hemos torpemente expuesto, presentó contra ella la reflexión que puede encontrarse hacia final del capítulo trece del *Leviatán*, sobre las precauciones que, incluso en la sociedad y bajo leyes y gobiernos estables, toman los hombres para evitar ser robados o agredidos, denunciando con ello la consideración en que tienen

[279] Ver, p. ej., Gauthier, "Economic Man and the Rational Reasoner", en Nichols, J.H. y Wright, C., *From Political Economy to Economics—and Back?*, San Francisco, Institute for Contemporary Studies Press, 1990, pp. 105-134; p. 107. Hay que reiterar inmediatamente, sin embargo, que la capacidad de representación semántica que permite la reflexión de segundo grado sobre las preferencias que daría lugar a las restricciones morales es —hasta donde sabemos— típica y exclusivamente humana.

a sus semejantes[280]. Nosotros podemos poner un ejemplo inspirado en el de Hobbes: pensemos en la función de los mercados en nuestras sociedades reales (y en cómo naturalmente surgen en el tráfico internacional). Aun en el marco de sociedades con gobiernos fuertes y estables, y entre ciudadanos educados para obedecer reglas razonables y dirigidas al bien común, donde la cooperación podría organizarse y planificarse racionalmente, resulta socialmente más ventajoso —como promedio— dejar una enorme parte de nuestra interacción al libre juego del auto-interés, incluso aunque haya que soportar los costes de un aparato legal y policial para evitar (o compensar) las constantes extralimitaciones e incumplimientos de los participantes en este juego. ¿No denuncia esto nuestra profunda creencia de que un hombre se mueve en primer lugar por su interés, y, muy en segundo término, por otras razones? Sin duda, el ejemplo es menos elocuente que el hobbesiano, pero tal vez nos permita apreciar hasta qué punto es razonable identificarnos con los individuos en la posición inicial, y falaz pretender que ellos no representan en absoluto la esencia de lo humano. Ellos representan, con bastante exactitud, lo que queda de racional si, dado un ciudadano medio de cualquiera de nuestras sociedades avanzadas, ponemos entre paréntesis su educación moral-religiosa, sus lazos familiares y de amistad y su temor al castigo.

No negamos que esto que queda es, en efecto, una abstracción irreal, pero mucho menos fantástica que otras hipótesis, igualmente abstractas, de rancio abolengo (o de repentino prestigio, como el "decisor tras el velo de ignorancia"). Se trata al fin y al cabo de un agente consciente de sus intereses, capacidades y deseos, y sabedor también de la posición que ocupa en la sociedad y la generación a que pertenece. Representa, en definitiva, una

---

[280] Nos referimos al conocido texto de Hobbes: "Puede parecer extraño a alguno que no haya valorado bien estas cosas que la naturaleza enfrente así a los hombres y les cree aptos para invadirse y destruirse unos a otros. Y puede consiguientemente que, no confiando en esta inferencia hecha a partir de las pasiones, desee quizá verla confirmada por la experiencia. Considere, entonces, que él mismo, cuando sale de viaje, se arma y procura ir bien acompañado; cuando se retira a dormir cierra con llave sus puertas; e incluso dentro de su casa cierra con llave sus cofres —y eso que sabe que hay leyes, y oficiales públicos armados, para vengar todas las injurias que le fueren hechas. Qué opinión tiene de sus semejantes cuando viaja armado, de sus conciudadanos cuando cierra con llave sus puertas, y de sus hijos y siervos cuando cierra con llave sus cofres, ¿No acusa él a la humanidad con sus acciones tanto como yo con mis palabras?".

Capítulo III

Hipótesis histórica: ¿Hay una tradición de contractualismo moral?

1. Presentación de nuestra hipótesis

En el capítulo anterior hemos expuesto los que podemos llamar "elementos del estado de naturaleza" —es decir, la construcción hipotética inicial de una teoría contractualista— y defendido su plausibilidad como premisas de un argumento moral dirigido al escéptico y al egoísta (y, *a fortiori*, a todos nosotros). Al caracterizar la racionalidad aludíamos, además, a la causa del conflicto en ese estado, a saber, la incompatibilidad entre la maximización individual ("lo racional" para cada uno) y la optimización colectiva ("lo racional" para todos). Esta contradicción se puede identificar metafóricamente con lo que ocurre en los mercados reales o imperfectamente competitivos: hay todo un conjunto de bienes que, pese a ser beneficiosos para todos, el mercado es incapaz de producir por sí mismo. Se llaman "bienes públicos" y se caracterizan porque, para ser producidos, exigen un tipo de comportamiento no directamente maximizador o competitivo, sino cooperativo. El símil de la producción de bienes públicos ayuda a entender que el único medio para escapar a la contradicción que se daría en la interacción natural entre agentes como los descritos en el capítulo anterior, es un acuerdo o pacto que, evitando que cada uno intente maximizar su utilidad dadas las acciones de los demás, les imponga —por su propio interés— la realización obligatoria de ciertas acciones (en general, la adopción de ciertas estrategias) sólo en vista de la utilidad de los otros, armonizando de este modo sus conductas y permitiéndoles alcanzar

óptimos sociales.

En los próximos capítulos seguiremos el razonamiento aquí anticipado, que concluye en la tesis de que un acuerdo es el único procedimiento que individuos independientes y auto-interesados aceptarían para establecer la base racional de la cooperación y la distribución de los beneficios de ésta. El argumento cae, como es evidente, en la tradición del contrato social, pero se diferencia de la mayor parte de las teorías contractualistas en que no está encaminado a justificar la obligación política ni a explicar el origen de la sociedad, las leyes o el estado. El argumento que aquí perseguiremos se endereza a justificar, ante cada individuo, la racionalidad de las reglas de la cooperación, esto es, las reglas que impulsan a dejar de lado el auto-interés cuando es mutuamente beneficioso hacerlo; unirse a las actividades cooperativas; y finalmente cumplir el "papel" que en esa actividad se comprometió a jugar, incluso aunque en el momento de cumplirlo sea individualmente costoso hacerlo. En la medida en que estas reglas de la cooperación, y su cumplimiento, se pueden identificar con la moralidad, el argumento que perseguimos se arroga la dignidad de "contractualismo moral", establecido sobre las firmes bases de la teoría de juegos y la negociación racional.

Muy bien puede decir Gauthier que la teoría moral, entendida como teoría de la decisión racional, es una empresa nueva, de la que él mismo se siente pionero[1], pero no puede negar que el uso teórico-político y filosófico de la idea de un contrato hipotético (o real) como medio de justificación del poder, el estado o la justicia, es tan antiguo como la filosofía misma, y aún más[2]. Y así lo reconoce en *MA*, al escribir que "históricamente, el contractualismo moral parece tener su origen entre los sofistas. Glaucón esboza una explicación contractualista del origen de la justicia en *La República* de Platón..."[3]. Es seguro que ambas afirmaciones pueden armonizarse, porque la Teoría de la Decisión Racional y el enfoque económico de la ética han permitido llevar la

---

[1] Cfr. Gauthier, D., "Morality, Rational Choice and Semantic Representation" (en Paul, E.F., *et al.* (eds.), *The New Social Contract*, Oxford, Blackwell, 1988, pp. 173-221), p. 173.

[2] La idea de que un contrato o acuerdo común de todo el pueblo, o del pueblo con el rey, o incluso del pueblo con Dios, es el origen de la unión social y de la obediencia debida a los monarcas se halla diseminada en distintos pasajes bíblicas sobre la historia del pueblo de Israel, una colección de los cuales puede encontrarse en el capítulo XI del *De Cive* hobbesiano, titulado "Lugares y ejemplos de la Sagrada Escritura sobre el derecho de gobernar, que parecen apoyar lo dicho".

[3] Pp. 9-10.

empresa contractualista tradicional a un grado de refinamiento que, no sólo ha perfeccionado su metodología, sino que ha ampliado también sus objetivos, convirtiéndola en una "empresa nueva". El sentido en que el neo-contractualismo moral representa una innovación filosófica, sólo podrá ser apreciado al final de nuestro estudio; sin embargo, una somera exploración de sus antecedentes históricos, ayudará a valorar la originalidad del proyecto aún antes de la exposición del argumento central.

Es definitivamente imposible intentar aquí una exposición exhaustiva de la historia del contractualismo político que, como hemos dicho, se remonta a los inicios de nuestra cultura y, en todo caso, tiene un representante en cada teórico político o ius-filósofo moderno. Por otro lado, el intento de reducir la moralidad a auto-interés o, con otras palabras, el intento de mostrar que "es racional ser moral" (lectura ética del contractualismo moral liberal) puede considerarse también el propósito perenne de la filosofía moral, con lo que tampoco por esa vía nos veríamos relevados de un estudio excesivamente largo para la modestia de nuestra capacidad y objetivos. No obstante, la naturaleza concreta del proyecto que afrontamos ("contractualismo moral") sí nos exime —si no lo hiciera igualmente el sentido común— de un estudio tan extenso. Porque si la "moral como parte de la Teoría de la Decisión" es una empresa nueva, el contractualismo moral tampoco ha sido, en general, una empresa demasiado frecuente, como intentaremos defender. Así, nuestro análisis, cuyo objetivo es simplemente situar en su contexto histórico la teoría de Gauthier, se centrará en intentar encontrar, entre el abultado y prestigioso número de los "contractualismos", algún verdadero antecedente del contractualismo moral tal como creemos que debe ser entendido.

Desde Glaucón, la filosofía política ha ofrecido, en efecto, múltiples modelos de contractualismo. Se han intentado variados sistemas de clasificación[4], algunos de los cuales pueden resultar útiles, pero no hemos encontrado

[4] La distinción más común es la establecida entre "contrato real o histórico" y "contrato hipotético". Algo más sofisticada es la distinción clásica (podemos encontrarla paradigmáticamente expuesta en la "Introducción" del *The Social Contract* de J.W. Gough [Oxford, Clarendon, 1957]) entre "contrato propio" o "pacto de asociación" (*Pactum Assotiationis, Gesellschaftsvertrag*) y "contrato impropio", "pacto de gobierno o de sujeción" (*Pactum subjectionis*) o *Herrschaftsvertrag*. Se trata de los dos tipos básicos de contrato social o político. Desde el punto de vista filosófico-epistemológico resulta más interesante, sin embargo, la distinción, sugerida por Jody S. Kraus (*The Limits of Hobbesian Contractarianism*, Nueva York, Cambridge U.P., 1993, pp. 2 y ss.), entre teorías analíticas y normativas del contrato: las primeras buscarían simplemente mostrar que la sociedad política puede entenderse en términos de un modelo teórico (la reconstrucción de un

ninguno que nos permita especificar el carácter de un contractualismo moral. Éste ha resultado esquivo a las conceptualizaciones tradicionales y resistente a adaptarse las taxonomías habituales. Hemos encontrado, por ejemplo, que el recurso de Gauthier a los argumentos de Glaucón como antecedente del contractualismo moral es equivocado, pues Glaucón explica convencional o contractualmente el origen de la sociedad y del respeto a la justicia, pero se puede observar —si se analiza con más detenimiento— que no ofrece una justificación racional individual de la obediencia a las reglas de la justicia, ni escapa al relativismo. Esto aleja sus tesis de los rasgos definitorios de un contractualismo moral propiamente dicho. Según nuestro punto de vista, el contractualismo moral debe identificarse con un tipo de teoría ética que prácticamente carece de antecedentes; ninguna de las teorías del contrato social que conocemos se adapta a la (creemos que razonable) definición que daremos del mismo.

Nos proponemos leer, por tanto, la historia del contractualismo a la luz de una hipótesis un tanto chocante, pero que intentaremos defender precisamente sobre la base de los datos proporcionados por el análisis de las teorías contractualistas históricas más relevantes.

Denominamos hipótesis, y no "tesis", a nuestra concepción del contractualismo moral justamente porque habrá de ser contrastada con la interpretación histórica y con el sentido del contractualismo moral contemporáneo, que aún no hemos analizado. Tanto aquella interpretación como éste análisis son, podría decirse, "maximalistas", esto es, suponen un punto de vista bastante radical en cuanto a la potencia filosófica del contractualismo moral y en cuanto a la pobreza de sus antecedentes putativos. Aceptamos que hemos agrandado un tanto la distancia a propósito, con el fin de enfatizar la originalidad y el alcance de la tarea del neo-contractualista liberal —cuya teoría consideramos, en muchos aspectos, más conectada con la filosofía trascendental que con el contractualismo político—, pero la eventual "distorsión" en que podamos incurrir no obsta para sostener y defender la corrección esencial de

---

contrato), a veces hipotético, o que los principios morales pueden entenderse como el producto de una elección especialmente cualificada; las segundas tratarían de demostrar, mediante una reconstrucción similar, que algo *debe* ser el caso, o está justificado o es legítimo. Básicamente esta misma división de Kraus es empleada por Gauthier cuando diferencia entre un convencionalismo descriptivo y otro normativo (Cfr. "Thomas Hobbes: Moral Theorist", en *Moral Dealing*, Ithaca, Cornell U.P., 1990, pp. 11-23, p. 16). Con todo, estas clasificaciones generales (sólo recogemos las más interesantes para nuestro propósito) no agotan, ni mucho menos, la rica taxonomía del contractualismo, parte de la cual iremos desgranando a lo largo del capítulo.

nuestro punto de vista. A ello dedicaremos el presente capítulo.

Sostendremos la hipótesis de que el contractualismo moral, propiamente entendido, supone un punto de partida escéptico sobre la moral, ocasionado, tal vez, por un estado de crisis de fundamentación moral[5]. Requiere, además, la afirmación del relativismo y subjetivismo de los valores y una concepción instrumental de la racionalidad. Desde tales condiciones y convicciones, la única justificación racional posible para lo que tradicionalmente denominamos moralidad consiste en demostrar que hay una relación íntima entre auto-interés y moralidad: que ser moral es verdaderamente beneficioso para cada uno. Pero, dado que, desde el punto de vista de la racionalidad individual, resulta palmario que no es beneficioso actuar "moralmente" en contextos competitivos, la moralidad únicamente puede demostrar su utilidad suponiendo que existan (o puedan crearse entre agentes racionales) ámbitos de cooperación. Tales ámbitos (la sociedad, por ejemplo) se erigen, a su vez, sobre las reglas de justicia, que son las que, al garantizar a cada individuo una porción aceptable del beneficio mutuo, lo motivan a entrar en esa empresa cooperativa. Así, la moral contractual se halla íntimamente unida a la justicia, es decir, al establecimiento y mantenimiento de instituciones imparciales y racionalmente legitimables. En el contexto de la cooperación y la justicia, la moralidad es el nombre de la mejor estrategia que puede adoptar un individuo perfectamente racional. Y a la vez se constituye en punto de referencia crítico contra aquellas instituciones injustificables desde el punto de vista de la razón individual: aquellas instituciones que no pudieran haber sido el resultado de un acuerdo racional entre agentes mutuamente desinteresados.

El principio de una moral por acuerdo es, pues, el objeto de un pacto encaminado a hacer posible la cooperación, esto es, el beneficio mutuo. El principio moral representa una condición necesaria para la unión social y, por ende, para la justicia y la obligación política. Pero a la vez, como el principio moral sólo tiene sentido en el ámbito de la cooperación (dentro de la sociedad), su eficacia (su obligatoriedad incluso) está supeditada a la justicia y el derecho, en la medida en que son los medios para el mantenimiento de aquella.

Una moral por acuerdo, cuya base racional estriba en que sea necesaria para alcanzar un beneficio mutuo, podría parecer, a primera vista, muy poca cosa según los cánones filosófico-morales tradicionales. Sin embargo,

---

[5] Podría aceptarse la representación del mismo ofrecida por la "inquietante sugerencia" del capítulo 1 de *Tras la virtud*, o el punto I del artículo "¿Por qué contractualismo?" de Gauthier (en *Doxa*, nº 6, 1989, pp. 19-38).

mostraremos que los trazos de instrumentalidad que la rodean no son óbice para considerarla como verdadera moralidad[6] y, sobre todo, que representa el único esquema racional de reglas y obligaciones que podría heredar ese nombre a partir de la situación y las creencias que expusimos arriba. En un estado de crisis de la fundamentación moral, la moral contractual ofrece la única alternativa racional al nihilismo (que adopta las formas de emotivismo, convencionalismo o escepticismo).

La clave de nuestra lectura histórica (y segunda parte de la hipótesis que defenderemos) será que este tipo de moralidad contractual esbozado —anticipo de la que habremos de construir en el capítulo siguiente sobre la base de las premisas expuestas en el anterior— sólo es posible desde una situación de la filosofía moral y social típica (aunque no exclusiva) de nuestro siglo. Por eso, los contractualismos anteriores (excepto tal vez Hobbes) y los intentos anteriores de reconciliar moralidad e interés, no pueden considerarse, estrictamente hablando, ejemplos de contractualismo moral (aunque sí supongan, evidentemente, un antecedente del contractualismo como método), ni siquiera en el caso de que expresamente se auto-denominen así, como acontece en la teoría de Glaucón sobre el origen de la justicia.

En nuestro siglo se ha insistido más seriamente que nunca en la tesis de que la moralidad no es más que un medio para auto-protegernos o beneficiarnos[7]. Desechadas las alusiones a la ley natural o divina, e incluso al imperativo categórico de la razón, sólo el interés de cada individuo parece una base suficientemente sólida sobre la cual establecer la misión esencial de la filosofía moral, que es justificar las demandas morales y explicar (y descubrir) la capacidad motivadora de éstas, si es que la tienen, o intentar aportarla mediante argumentos si carecen de ella. Y el modo en que el auto-interés da paso a la

---

[6] En este sentido, es iluminador el comentario de Scanlon al final de su artículo "Contractualism and Utilitarianism" (en Sen A. y Williams, B. [eds.], *Utilitarianism and Beyond*, Cambridge U.P., 1982, pp. 103-128). Allí diferencia Scanlon entre la consideración instrumental de la moralidad como una mera estrategia al servicio de la protección mutua, y la visión contractualista, según la cual, nuestro interés en protegernos influye en la decisión sobre lo que es racional pactar y dota, así, de *contenido* a la moralidad. La moralidad acaba promocionando nuestros intereses. Pero, si tenemos en cuenta que lo hace tras un acuerdo unánime sobre qué es razonable hacer en vista de nuestros intereses, y que el deseo de llegar a ese acuerdo posible ("de ser capaces de justificar razonablemente nuestras acciones ante los demás" en términos de Scanlon) es anterior al pacto mismo, la moralidad puede verse como "menos instrumental".

[7] Cfr. Scanlon, "Contractualism and Utilitarianism", cit., p. 128.

moralidad es el contractualismo, un método filosófico para tiempos de crisis.

Ante la crisis de las fuentes tradicionales de la normatividad moral, el expediente del contrato —que sirvió desde el siglo XVII para suplir la pérdida o deterioro de las antiguas fuentes de la normatividad política y social— es empleado en nuestro tiempo como alternativa ética. Si la idea del contrato entre hombres libres sirvió para legitimar el poder político cuando la legitimación tradicional hizo crisis, la nueva idea del contrato entre agentes auto-interesados (a-morales) debe servir para "derivar principios de justicia social"[8] cuando cualquier fundamento "fuerte" de la moralidad ha sido igualmente criticado.

Tal vez el único antecedente aproximado de este uso radical del contrato en tareas de fundamentación moral esté representado por Thomas Hobbes —o, mejor dicho, por una lectura posible de la teoría hobbesiana— debido a que una conjunción de circunstancias históricas y personales le condujeron a adoptar unas premisas (escepticismo, egoísmo, subjetivismo axiológico) muy semejantes a las que se nos imponen en nuestros días. Trataremos de justificar esta afirmación más abajo, donde precisaremos sus límites.

En general, sostendremos que, a excepción de una peculiar lectura de Hobbes, según la cual podría ser calificado de "contractualista moral frustrado", la llamada historia del contractualismo agrupa un conjunto de doctrinas morales, políticas y filosófico-jurídicas que pueden clasificarse, al menos en dos grandes grupos: convencionalistas y contractualistas. Afirmaremos que existe un contractualismo político y jurídico, y también un convencionalismo político y jurídico; pero en lo que se refiere a la moral, sólo se puede hablar de convencionalismo moral y no (hasta nuestro siglo) de verdadero contractualismo moral.

No hace falta insistir en que se trata de una propuesta tentativa, una conclusión a la que hemos llegado no sin atravesar dudas y perplejidades acerca del verdadero sentido de un contractualismo moral, y que depende enteramente de nuestra comprensión del mismo. A fin de aclarar algo esta comprensión y

---

[8] La frase es de Gregory Kavka, y se inscribe en el siguiente párrafo, del cual extraemos la idea del nuestro: "La versión hobbesiana de la teoría del contrato hipotético difiere de las de escritores contemporáneos como John Rawls y David Gauthier en, al menos, un aspecto fundamental. Mientras estos escritores tratan de derivar principios de justicia social del contrato hipotético, Hobbes (y la teoría hobbesiana) usa este expediente para un propósito más modesto: identificar las condiciones que debe satisfacer un sistema político o Estado para que sus habitantes o ciudadanos estén obligados a obedecer sus reglas." (en *Hobbesian Moral and Political Theory*, Princeton, Princeton U.P., 1986, p. 182).

de allanar el terreno para la defensa de nuestra hipótesis histórica, será necesario llevar a cabo una distinción y varios comentarios previos. Debemos distinguir entre convencionalismo y contractualismo. Tal distinción será la clave y fundamento de nuestra interpretación histórica, y está apoyada por definiciones derivadas del análisis textual. No será, por así decir, una distinción "descriptiva", o una mera clasificación, sino una distinción "normativa", que trata de establecer qué condiciones tiene que tener una teoría para poder ser denominada propiamente "contractualista". Establecida esta distinción, repasaremos, a su luz, el contractualismo histórico, con especial atención al caso de Hobbes, para concluir defendiendo nuestra perspectiva histórica del contractualismo moral liberal, que lo inscribe en una limitadísima tradición.

2. Diferencia entre convencionalismo y contractualismo

La filosofía política tiene entre sus objetivos responder a las preguntas por la legitimidad de los gobiernos, la naturaleza y formas de la justicia, el carácter de las leyes, la justificación de la obligación política, etc. Todas estas cuestiones pueden difícilmente separarse del contexto en que se plantean. El filósofo político está frecuentemente acuciado por la necesidad de legitimar o criticar un poder monárquico, tiránico, democrático o republicano concreto. La historia política condiciona la reflexión y supone que cada teoría contiene muchos matices que es imposible considerar. Debido a esta inmediata relación con la praxis, las teorías políticas no suelen preocuparse por establecer distinciones metodológicas ni meta-teóricas, sino que emplean los mejores o más apropiados recursos para sus respectivos argumentos. La neta distinción entre convencionalismo y contractualismo introduce, pues, una rigidez que no existe en la filosofía práctica. Hablaremos, consiguientemente, de tipos ideales en los que probablemente ninguna teoría puede inscribirse sin reservas[9]. No obstante, servirán para situar los dos extremos de un espectro en algún punto del cual toda teoría pretendidamente contractualista tiene su lugar. A la vez, permitirán distinguir qué parte de una teoría puede considerarse propiamente contractualista y cuál no, etc.

a) La confusión de los términos.-

Antes de iniciar el análisis que nos conducirá a establecer las semejanzas y diferencias entre convencionalismo y contractualismo, debemos dejar constancia de la confusión e imprecisión con que se usan estos términos. De hecho, el anárquico uso de ambos términos: "contrato" y "convención" (y sus sinónimos), en las obras de la mayor parte de filósofos, nos ha —más que sugerido— impuesto el contenido de este epígrafe. La general desatención en

[9] En especial, las teorías más antiguas, mucho más difusas y menos cuidadosas con el uso de terminología y método que las modernas.

este aspecto lleva a agrupar bajo una misma denominación a teorías que difieren fundamentalmente por sus objetivos, su método, sus premisas filosóficas y su alcance. Sin ánimo de exhaustividad, se pueden recoger algunos de los variados, y no siempre bien definidos, sentidos en que se han empleado estos conceptos.

Ya apuntamos arriba una útil división general entre contractualismos analíticos y normativos, según se enderezasen a explicar cómo puede entenderse el Estado o la justicia en términos de un pacto o elección originarios, o más bien a justificar la legitimidad o racionalidad de esas mismas instituciones y prácticas. En la mayoría de los casos, esta distinción no es evidente, aunque habremos de aludir a ella con frecuencia por su utilidad para nuestro análisis.

La diferencia entre pacto de unión (por el que un grupo de individuos se agrupan para formar un pueblo) y pacto de sujeción (por el que un pueblo acepta someterse a un príncipe y éste se obliga a regirlo de acuerdo a ciertas leyes); así como la diferencia entre teorías históricas y teorías hipotéticas del contrato, no suelen ocasionar confusiones, aunque tampoco están completamente exentas de ellas[10].

Hume, por su parte, habla de cuatro tipos de contractualismo, a saber: contractualismo *original* (tesis de que el origen de la sociedad y sus instituciones básicas se halla en una convención pactada), contractualismo *explícito* (que defiende la legitimidad del gobierno apelando a un acuerdo real, bien entre los ciudadanos, bien entre éstos y sus gobernantes), contractualismo *tácito* (que sostiene que la aceptación de las ventajas derivadas de vivir en un Estado y al amparo de sus leyes, implica el asentimiento al sistema político que lo mantiene; según esta teoría, solo la rebelión o el exilio valen como pruebas en contra del acuerdo legitimador), y contractualismo *hipotético* (según el cual los sistemas de propiedad y gobierno son legitimados en la medida en que podrían ser acordados por personas racionales en cierta situación ideal).

El contractualismo original alude al uso meramente histórico-explicativo del contrato; el segundo y tercer tipo son variantes de la teoría del consentimiento político o constitucional, muy usada a lo largo de la Edad Media y moderna como explicación del origen y legitimidad del poder real. Estas tres versiones del contractualismo no pueden ser consideradas auténticas teorías filosófico-políticas. El contractualismo hipotético, por el contrario, sí se acerca

[10] Así, por ejemplo, Rousseau reprueba sutilmente, en el capítulo V de *El contrato social*, el uso que Grocio hace de la idea de *pactum subjectionis* y su olvido de que el *pactum unionis* ha de ser lógicamente anterior a aquél. Y, en general, la tradición escolástica trató equívocamente el *pactum subjectionis* como pacto fundante de la comunidad política.

al contractualismo por antonomasia, una teoría filosófica dirigida a la legitimación racional del Estado.

La alusión a las teorías medievales sobre el origen del poder regio, y su oposición a las teorías modernas, nos invita a notar otro equívoco. Se trata de que se ha denominado contractualistas por igual a teorías que entienden el pacto como mero consentimiento (incluso tácito, como veíamos) y a teorías cuyo argumento depende de un "verdadero" acuerdo o contrato entre iguales (hombres libres) del que se derivan obligaciones mutuas. No se trata de una diferencia nominal: el contrato es un negocio cuyo origen y esencia está en la voluntad de las partes. Las partes, en uso de su libertad, *crean* una nueva estructura de derechos y obligaciones entre ellas. Al hablar de consentimiento esto no es necesariamente así: J.W. Gough nos recuerda que el escolástico Luis de Molina comparó el Estado con el matrimonio, que necesita del *consentimiento* de las partes para existir, pero no es creado o instituido por ese consentimiento[11]. Con esta analogía se evidencia cómo la teoría del consentimiento, aparentemente contractualista, sigue anclada, en muchos aspectos, a una concepción política aristotélica: el Estado requiere necesariamente del consentimiento de los ciudadanos para su existencia legítima, pero, como tal institución, es de origen divino, a través de la naturaleza, que hace a los hombres sociables[12]. Esta concepción ha desaparecido entre autores que, como Hobbes o Spinoza, parten de una naturaleza humana a-social y hacen depender tanto la legitimidad del Estado como su origen de un acuerdo fundante entre iguales. Pese al significativo antecedente neo-escolástico, habremos de reservar el término "contractualismo" para estos últimos autores.

El convencionalismo también se ha entendido en varios sentidos. No obstante, hay mayor acuerdo sobre la definición de una "convención". Como

[11] Cfr. Gough, J.W., *The Social Contract*, Oxford, Clarendon, 1957, p. 68.

[12] Debemos apuntar inmediatamente que, pese a lo que pueda dar a entender el tono de nuestro argumento, no consideramos que la pervivencia de restos aristotélico-tomistas sea un demérito de la neo-escolástica española. Todo lo contrario, esta escuela puso las bases (con sus teorías sobre el derecho de resistencia, los derechos naturales individuales, etc.) que habrían de conducir a un individualismo y contractualismo radicales allí donde el desembarazo definitivo de la tradición fue más fácil, y puede considerarse antecedente directo de aquellos otros autores que —como Locke— se mantuvieron más fieles a la idea de una Ley Natural de origen divino. Cfr. más abajo, punto 2. b), donde se añadirá algo sobre esto.

nos recuerda Harman[13], Hume definió con bastante exactitud una situación convencional como aquella en la que todos se adhieren a un principio de modo que otros (se entiende, recién llegados) encontrarán razonable adherirse igualmente. En una situación convencional, cada uno espera la adhesión de los demás en tanto él mismo se mantiene fiel a la convención, de forma que lo que la mantiene estable es simplemente el hecho de la adhesión unánime o, al menos, muy mayoritaria[14]. Las convenciones son, por tanto, regularidades arbitrarias o contingentes, cuya normatividad se asienta en su eventual utilidad *de facto*. De ahí deriva precisamente un segundo sentido en que se usa el término. Porque autores como Rousseau, heredando la oposición clásica entre *physis* y *nomos*, hablan de "convención" para referirse a las instituciones no-naturales, mantenidas por la mera voluntad y conveniencia humanas[15]. Este segundo sentido se centra en el contenido (y no tanto en la forma) de las convenciones, y enfatiza el carácter arbitrario de las mismas[16].

---

[13] Cfr. Harman, G., *The Nature of Morality*, Nueva York, Oxford University Press, 1977, p. 103.

[14] Este mismo sentido de "convención" es formalmente definido por Gauthier como sigue: "Definiré una *convención* como una regularidad *R* en el comportamiento de las personas *P* en situaciones *S*, tal que parte de la razón que la mayoría de estas personas tienen para actuar conforme a *R* en *S* es que es de dominio público entre *P* que la mayoría actúan conforme a *R* en *S*, y que la mayoría espera que la mayoría (de los demás) actúe conforme a *R* en *S*." ("Thomas Hobbes: Moral Theorist", en *Moral Dealing*, cit., pp. 11-23; p. 16).

[15] Cfr. *El contrato social*, cap. II.

[16] No nos detendremos en otros posibles sentidos del concepto "convención". Sí queremos comentar, sin embargo, que la dicotomía *physis/nomos*, que sirve como marco categorial en el caso de Rousseau, ha sido criticada por un sector de la tradición liberal, al poner el énfasis en el nacimiento de las convenciones. Según esta tradición (cuyo perspicuo representante actual es F. Hayek, pero que se remonta a Mandeville, Ferguson y Smith) la distinción entre lo "natural" (*physei*) y lo "artificial" (*nómo* o *thései*) "interpretada como una alternativa excluyente, no sólo es ambigua, sino decididamente falsa; como terminaron por verlo claramente los filósofos sociales escoceses del siglo XVIII (...) una gran parte de las formaciones sociales, aunque son el resultado de la acción humana, no lo son de un designio humano" (Hayek, *New Studies in Philosohy, Politics, Economics and the History of Ideas*, Londres, Routledge and Kegan Paul, 1978, p. 4-5). Esta tradición considera que muchas "formaciones sociales" (entre ellas los sistemas normativos) son "artificiales" en el sentido de ser un resultado de acciones humanas, mientras los acontecimientos naturales son ajenos a ellas; pero no son literalmente *artefactos* deliberadamente creados con vistas a un fin.

La idea que guía la tesis de Hayek es que la acción humana puede producir "espontáneamente" órdenes normativos mediante procesos de coordinación (Cfr., sobre la coordinación, más abajo, cap. 4, punto 2.c). La interacción coordinativa se caracteriza por su estabilidad, ya que surge porque representa un beneficio inmediato para los agentes implicados. Es dudoso que mecanismos de coordinación espontánea sean capaces de dar lugar a estructuras normativas o de

A estos dos sentidos principales de "convención" —que consideramos, además de análogos y complementarios, acertados— habrá que añadir todos los sentidos espurios, que equivocadamente identifican "convención" y "convencionalismo" con alguno de los varios significados de "contrato" y "contractualismo" que acabamos de reseñar, o con algún otro[17].

De hecho, el mayor problema que se presenta al intentar definir el contractualismo es que ni siquiera este simple desbroce conceptual que acabamos de realizar a modo de ejemplo es tenido presente, y así convencionalismo y contractualismo se mezclan y confunden. Desde luego, ambos tienen puntos en común (en especial sus premisas) que hacen que teorías convencionalistas y contractualistas empleen un lenguaje muy semejante. Pero si sus orígenes y forma pueden ser similares y dar lugar a confusión, no lo es así su contenido filosófico-político. Sus diferencias son mucho mayores y más profundas que sus analogías. Comenzaremos, sin embargo, por estudiar sus puntos en común y explicar, de paso, las razones posibles de la confusión.

---

interacción complejas (aunque el mercado sería, según Adam Ferguson, un argumento contra esta duda; cfr., p. ej., su *An Essay on the History of Civil Society*, Edimburgo, 1767, p. 187); sin embargo, los defensores de este punto de vista argüirán que entre las situaciones (relativamente simples) de coordinación espontánea y aquellas otras en que existe un marco normativo expresamente diseñado para promover y/o asegurar la cooperación, puede darse una gama indefinida de situaciones intermedias como serían las costumbres o las convenciones.

En este sentido, puede verse Thomas C. Schelling, *The Strategy of Conflict*, Cambridge, Harvard U.P., 1960 (trad. española: *La estrategia del conflicto*, Madrid, Tecnos, 1964), que es uno de los primeros estudios sistemáticos de la génesis de las convenciones como solución a problemas de coordinación; y David K. Lewis, *Convention. A Philosophical Study*, Cambridge, Harvard U.P., 1969. Entre los seguidores de lo que Elster denomina "el programa de Hayek" (cfr.*The Cement of Society*, Cambridge, Cambridge U.P., 1989, p. 250), cuyo propósito es investigar cómo es posible un orden espontáneo en la sociedad, se contarían también Edna Ullman Margalit, *The Emergence of Norms*, Oxford, Clarendon, 1977; Russell Hardin, *Collective Action*, Baltimore, John Hopkins U.P., 1982; Michael Taylor, *The Possibility of Cooperation*, Cambridge, Cambridge U.P., 1987; Robert Axelrod, *La evolución de la cooperación*, Madrid, Alianza, 1986.

[17] Es muy común, por ejemplo, el error que consiste en denominar "convencionales" a todas las normas que no cabe calificar como "naturales" o de origen divino, sea cual sea su justificación (propiamente convencional o contractualista). Creemos que Gauthier cae en este error al comentar la teoría de Hobbes en "Thomas Hobbes: Moral Theorist" (en *Moral Dealing*, cit, pp. 11-23), cfr. p. ej. p. 17.

b) Puntos de partida comunes.-

El contractualismo y el convencionalismo son teorías (o tipos de teorías) filosófico-políticas o ius-filosóficas para tiempos de crisis[18]. Se trata de modelos explicativos (incluso normativos) que exigen mínimos compromisos metafísicos previos y, de esta forma, es natural que surjan cuando la confianza en las certezas compartidas se ha resquebrajado. En esos casos, el recurso al pacto o la convención parece el único fundamento —muchas veces contingente e inestable— al que cabe recurrir de buena fe.

Aunque, como veremos inmediatamente (y ya hemos anticipado), los paradigmas contractualista y convencionalista suponen enfoques filosóficos de muy distinto alcance, ambos se distinguen muy claramente de todos los demás modelos de explicación o justificación moral. Para aclarar esta diferencia, que acota nuestro campo al diálogo entre convención y contrato eliminando la referencia a otros paradigmas de la filosofía práctica, tal vez sea pertinente recurrir a un ejemplo:

Pondremos por caso la institución social de la esclavitud. Como toda institución, implica un conjunto de reglas cuyo cumplimiento implica y causa obligaciones y deberes mutuos entre una serie de individuos —deberes que frecuentemente chocan con los intereses o deseos inmediatos de los obligados. Además, la institución como tal, genera un beneficio o coste neto para la sociedad en su conjunto.

De entre los modos de justificar racionalmente esta institución, destacaremos tres: un modo utilitarista, otro convencionalista y un tercero al que denominaremos "naturalista". Según el argumento utilitarista, la institución *debe* mantenerse siempre que el beneficio total que produzca exceda su coste, es decir, siempre que la suma de las utilidades que implique para distintos agentes exceda la pérdida de utilidad conjunta de quienes se vean perjudicados (los esclavos y/o quienes simpaticen con ellos). El argumento "naturalista" afirmaría que "hay esclavos por naturaleza", de modo que la institución *debe* mantenerse porque refleja una distinción natural entre los hombres y, por lo tanto, está legitimada por la naturaleza (es *buena* o, al menos, está permitida por naturaleza). La explicación convencional (o, a estos efectos, contractualista)

[18] El gran momento del contractualismo político clásico corresponde a la crisis producida por la aparición en la escena filosófica del Sujeto, la Razón y la Ciencia modernos. El convencionalismo aparece, como una idea recurrente desde la antigüedad griega (ejemplificado en la sofística y en algunas escuelas helenísticas), siempre que la crisis afecta a los fundamentos de las creencias (en sentido orteguiano) prácticas compartidas.

aceptaría la institución como razonable sólo si se demuestra que es (o podría ser) el fruto de un convenio voluntario y no engañoso (histórico o ideal) entre todas las partes implicadas en la misma.

Hay que observar que el fundamento de legitimidad de la institución, según el convencionalismo, reside en la racionalidad individual de los mismos implicados (en su decisión conjunta), mientras que en los primeros casos, ese fundamento reside en algo externo a la institución misma: la utilidad o "la naturaleza humana". Por así decir, la justificación convencional —a diferencia de las otras dos— se cierra sobre sí misma, sin necesidad de puntales metafísicos externos[19]. Pero no hay que olvidar que esta aparente autonomía del modelo convencional de justificación se basa en una radical afirmación de la capacidad individual y un insobornable respeto a los individuos en cuanto tales. Por último, hay que añadir que el carácter distintivo de la justificación convencional-contractual destaca si se compara la ligazón que se establece, en el naturalismo y el utilitarismo, entre ciertos postulados metafísicos[20] y las conclusiones normativas, con el carácter básicamente contingente de la justificación convencional. En ésta última, el procedimiento argumentativo (o histórico) no prejuzga el resultado: tan justificable puede resultar la institución de la esclavitud como su negación —aunque es presumible que no se aceptará tan poco equitativo sistema sobre la base de un pacto. Pero como tal convención, ésta depende de datos contingentes concernientes a la naturaleza humana y a la situación concreta[21].

---

[19] Desde nuestro punto de vista, ésta es la gran ventaja del modelo contractualista-convencionalista,l a que lo hace tan interesante en tiempos de crisis de creencias. La justificación naturalista puede ser muy sólida, pero traslada el centro de gravedad filosófico a la metafísica, pues requeriría —por seguir con el ejemplo de la esclavitud— *demostrar* que la naturaleza humana es efectivamente como se pretende que es.

[20] En el caso del utilitarismo, podría decirse que se trata de postulados metafísicos y pseudocientíficos, como los que aseguran que es posible sumar las utilidades de los miembros de una sociedad a efecto de conocer cuál es la "utilidad social" o "pública" de cierto comportamiento o institución.

[21] El ejemplo clásico de convención es el código de circulación. Sus mandatos son enteramente contingentes (tan es así que en una parte del mundo se circula por la derecha y en otra por la izquierda). Se justifican únicamente porque han sido objeto de una convención o acuerdo *y todos los respetan*. Una vez establecido esto, sería absurdo intentar construir argumentos ulteriores sobre la preeminencia "natural" de uno de los lados de la calzada. El *quid* de la justificación convencional está precisamente en que sólo se trata de eso, de una convención. Y frente a la fácil objeción de que este modelo de reflexión filosófica siempre justificará lo mayoritariamente observado y será, por tanto, conservador, sugerimos pensar si es fácil imaginar qué tipo de acuerdo o práctica común puede dar lugar a una institución como la esclavitud. Aparecerá claramente que

La gran diferencia entre convención y fundamentación natural es así evidente: la segunda requiere aceptar un compromiso metafísico, la primera no lo exige en absoluto o, al menos, no en tan alto grado. La segunda está prohibida fuera de una concepción global del mundo (en la que se inscriben las convicciones metafísicas en las que se apoya); la primera es posible y aceptable incluso para sujetos con distintas cosmovisiones[22].

Hemos afirmado que el paradigma convencionalista-contractualista se adapta a períodos de crisis en las creencias. Sin embargo, "tiempos de crisis" no significa "tiempos en que no se cree en nada"; la imagen más bien sería la de momentos en que, ante el derrumbe de antiguas certezas, la filosofía se refugia en ciertos "valores seguros": la dignidad de cada agente individual, la búsqueda privada de la felicidad a través de lo más útil o bueno para cada uno, etc. Estos conceptos, y otros similares, perfilan los contornos del clima en que surgen convencionalismo y contractualismo, y les proveen de una base —la única posible en tales momentos— para sus respectivos razonamientos. Así, aunque el alcance de ambos modelos será, después, muy diferente, tienen en común, el menos, el punto de partida.

En primer lugar, tanto convencionalismo como contractualismo se basan en una premisa individualista. Ya mostramos en el capítulo anterior que cierto grado de individualismo es necesario para poder hablar de contractualismo. Ello así en cualquier ejemplo de estas teorías. Contemporáneamente, el individualismo adopta la forma de "metodológico"; en las versiones clásicas se trata de una más o menos difusa confianza o defensa de la dignidad del individuo y de la capacidad creadora del sujeto individual. En definitiva, el punto de anclaje del convencionalismo y del contractualismo es el individuo o, mejor, dicho, los individuos. La diferencia entre uno y otro estriba en que el convencionalismo se fija en el hombre como ser-empírico, mientras que el contractualismo lo enfoca como sujeto transcendental (si se nos permite esta terminología kantiana). El convencionalismo se centra en el dato de que "estos" individuos arrojados en el mundo son la única medida de todas las cosas; el contractualista

---

el convencionalismo, lejos de legitimar cualquier práctica, es sumamente selectivo en el tipo de instituciones o reglas que permite justificar.

[22] Por esto afirmamos que una fundamentación "natural" es propia de momentos y filosofías "fuertes" —como el período arcaico en Grecia y las filosofías platónica o aristotélica— mientras que los expedientes convencionalistas son comunes cuando predomina el escepticismo y el relativismo, como sucedía entre los sofistas.

encuentra que, aunque se hayan perdido los demás fundamentos, *hay* una capacidad creadora en el sujeto, que se expresa a través de su racionalidad, y que esa capacidad permite reconstruir parte de los fundamentos perdidos[23].

Un segundo punto común es la concepción de la justicia y la moral. Ambos modelos parten de la idea de que la justicia y la moralidad son instituciones humanas que sólo tienen sentido en el marco de las sociedades reales o posibles. En palabras de Hobbes, el ámbito de la moralidad y la justicia es la "mutua conversación y sociedad de la humanidad"[24] —en un sentido análogo Rawls escribirá que "la justicia es la primera virtud de las instituciones sociales"[25]. Esto supone que ni la justicia ni la moral, ni la virtud, son un dato previo: lo justo, lo virtuoso, incluso lo bueno objetivo, dependen del acuerdo, del pacto o de la convención.

Más adelante matizaremos mucho esta concepción, especialmente por lo que se refiere al contractualismo, pero, por ahora, queremos expresar esta novedad que ambas doctrinas representan: tanto convencionalismo como contractualismo rechazarían la idea de que hay una especie de regla absoluta de justicia, o una virtud personal, anterior a la idea de una sociedad enderezada al beneficio mutuo.

Relacionado con esta concepción de la justicia y la moralidad está el reconocimiento de su carácter utilitario. El convencionalista reconocerá que, tanto el establecimiento de principios como la conformidad a los mismos merece la pena sólo si de ello se deriva un beneficio para todos y cada uno. El contractualista, aunque con una comprensión más profunda, también inicia su argumento desde bases semejantes.

Por último, también la visión de la motivación humana es compartida. El auto-interés es considerado el impulso principal que mueve a los hombres en sus acciones. En este punto, la diferencia está en que el contractualismo

---

[23] Con ello, el contractualista explica indirectamente la causa de la pérdida de los antiguos fundamentos, cosa que el convencionalismo ni intenta. El contractualismo afirma haber encontrado la única base racional firma para una normatividad universal. Ante su hallazgo, es posible juzgar que los paradigmas anteriores (religioso, naturalista, etc.) son inútiles (por innecesarios) como bases para la justificación de las normas, por lo que *debían* ser abandonados.

[24] Hobbes, *Leviathan*, cap. XV.

[25] Rawls, J., *Teoría de la justicia*, cit., p. 19. Esta afirmación de Rawls no excluye, en principio que la justicia pudiera tener otros ámbitos, pero sí creemos que restringe deliberadamente el ámbito de la justicia política en un sentido análogo al hobbesiano, e incluso más preciso.

buscará una relación *necesaria* entre auto-interés racional y reglas morales o estructuras políticas, no así el convencionalismo.

Como consecuencia de estas premisas comunes, contractualismo y convencionalismo ofrecen frecuentemente un rostro parecido: surgen como alternativa a teorías políticas de origen religioso o base metafísica, adoptan como única premisa la libertad individual que conduce a cada agente a perseguir su propio interés empleando los medios más apropiados, creen poder hallar, así, el origen o fundamento de la sociedad y el Estado en un acuerdo mutuamente ventajoso entre individuos independientes, con la consecuencia de que la justicia aparece como una creación consensual post-contractual y la obligación política como una convención útil mientras todos, o una significativa mayoría, se acomoden a ella.

Lo que debemos analizar a continuación es si esta apariencia esconde un parentesco cierto, o una semejanza casual. Para ello, nos centraremos, con mayor concreción, en algunos criterios relevantes sobre los que podría establecerse la diferencia entre convencionalismo y contractualismo.

c) Bases para la distinción.-

Tres criterios relevantes pondrán inmediatamente de manifiesto las diferencias esenciales entre teorías convencionalistas y contractualistas. Estos tres criterios son, primero, su relación con el expediente teórico del contrato; segundo, su carácter normativo; y, tercero, su racionalismo.

1.- Un primer criterio para la distinción será la aparición o no de un contrato en el argumento. Se trata de un criterio formal, no es definitivo, pero esconde un significado más que formal que esperamos aclarar.

El problema de si las teorías llamadas del "contrato social" envuelven un verdadero contrato o simplemente ejemplifican una situación convencional tal como la definimos arriba, ha sido suscitado por Jean Hampton en su libro *Hobbes and the Social Contract Tradition*. Allí argumenta que las teorías del contrato —en especial la de Hobbes, que ella analiza— suponen "acuerdos auto-interesados", pero no contratos, y que "los acuerdos auto-interesados difieren de los contratos en que son coordinaciones de intenciones para actuar que las partes mantienen *únicamente por razones de auto-interés*, mientras los contratos son negocios promisorios que introducen incentivos morales que, bien

complementan, bien *reemplazan* las motivaciones auto-interesadas de cada parte"[26]. La diferencia que Hampton observa entre lo que llama "acuerdos auto-interesados" y "contratos" es susceptible de ser identificada con la diferencia que, en la teoría de juegos, se establece entre "juegos de coordinación" y "juegos de cooperación". Los primeros se refieren a aquellos casos en que el interés particular de cada jugador se ve satisfecho en mayor grado si efectivamente contribuye a la actividad conjunta. El ejemplo clásico es el caso de dos personas en una barca de remos: cada uno tiene interés en remar si el otro rema, pues así la barca avanza, pero no si el otro no rema, pues entonces lo único que consigue es hacer girar la barca en círculo. En estas situaciones las actividades conjuntas se inician y mantienen sin necesidad de coacción, debido sólo al interés particular de cada agente. El paso de la situación en que ninguno rema a la situación en que ambos reman no sólo es beneficioso para *todos*, sino también para *cada uno*, con lo que es un cambio factible y estable, fundado en la prudencia individual. Los juegos de cooperación, por el contrario, representan situaciones en que la obtención de un resultado beneficioso (en términos de utilidad) para todos supone, para alguno o todos los jugadores, renunciar a la estrategia directamente maximizadora. El dilema del prisionero podría valer como ejemplo. En este juego, la obtención del resultado óptimo exige que cada uno renuncie a su estrategia maximizadora. Esta renuncia es perfectamente racional *ex ante*, pues cualquiera puede ver, y prefiere, los beneficios del acuerdo mutuo, pero sigue siendo irracional (en términos de maximización) *ex post*, cuando el pacto ha de ser cumplido. El único modo de lograr el resultado óptimo es realizar un verdadero contrato, esto es, una promesa tal que cambie la situación inicial del juego, bien introduciendo coacciones, bien produciendo un cambio psicológico en los jugadores, tal como argumenta Parfit[27]. La cooperación se diferencia así de la mera coordinación en que es inestable y necesita ser, como dice Hampton, complementada o incentivada con motivaciones morales o políticas.

Pero, entonces, el argumento de Hampton viene a ser que el problema que surge en el estado de naturaleza es un problema de coordinación, que cabe resolver sin recurso a un contrato en sentido estricto. Tal argumento resulta claramente implausible. Es evidente que el cumplimiento del pacto social

---

[26] Hampton, J., *Hobbes and the Social Contract Tradition*, Cambridge, Cambridge University Press, 1986, p. 147.

[27] Cfr. Parfit, D., "Prudencia, moralidad y el dilema del prisionero", *Diálogo Filosófico*, 1989, pp. 4-30; esp. pp. 7-8.

supone la realización, por parte de cada agente, de acciones no directamente maximizadoras que demandan, para la seguridad de todos los demás (y de sí mismo, dado su propio interés en que el pacto se cumpla) la introducción de incentivos en forma de sanciones (sean externas, como las penas legales, o internas, como el sentimiento de culpa). Este razonamiento lleva a Gauthier a concluir que la institución del soberano revela, en la teoría de Hobbes, el carácter esencialmente contractual (y no coordinativo) del acuerdo entre los sujetos[28].

Este argumento referido a Hobbes se puede generalizar afirmando que las teorías convencionalistas son aquellas que entienden que la situación original establece unas condiciones similares a las que dan lugar a juegos de coordinación o semi-coordinación[29]. En esta situación, el pacto es innecesario; como acertadamente dice Gauthier, "dado que una convención dominante y estable no necesita ni negociación ni pacto, no deja lugar a contrato alguno"[30]. Por el contrario, las teorías contractualistas ven la solución al conflicto natural como un problema de cooperación, en el que es necesario un pacto apoyado por sanciones, es decir, un contrato propiamente dicho.

Con todo, este criterio para diferenciar teorías convencionalistas y contractualistas no es definitivo. Existen explicaciones convencionales de la justicia que no acaban de adaptarse al mismo. Por ejemplo, Trasímaco propone que la justicia es lo que conviene a los poderosos[31]. Esto es, sin duda, una convención entre los poderosos. Una acción coordinada que les beneficia a todos y cada uno de ellos, sin embargo, en cuanto atañe a otros (los no

---

[28] Cfr. Gauthier, D., "Taming Leviathan", *Philosophy and Public Affaires*, 16 (1987), pp. 280-198, esp. pp. 296-7; y "Hobbes's Social Contract", en Rogers, G.A. y Ryan, A., *Perspectives on Thomas Hobbes*, Oxford, Clarendon, 1988, pp. 125-152; esp. p. 138.

[29] Un juego de coordinación puro es el ejemplificado por los dos remeros: en él, la elección de ambos jugadores converge en una acción conjunta que resulta ser estable y dominante. Los juegos de coordinación mixtos o semi-coordinación, como los hemos denominado, se suelen ejemplificar con el juego conocido como "batalla de los sexos", en el cual, los jugadores son los miembros de una pareja formada por un hombre y una mujer. Sus preferencias convergen en el hecho de que ambos prefieren pasar una velada juntos, pero difieren en que el hombre prefiere ir al fútbol y la mujer prefiere ir al teatro. Así, el orden de preferencias de la mujer será el siguiente: 1°.- ir juntos al teatro, 2°.- ir juntos al fútbol, 3° y 4°.- ir sola al teatro o al fútbol. Las preferencias del varón son semejantes, salvo que los dos primeros miembros se invierten. en este juego la solución dependerá del peso relativo de las preferencias de cada miembro pero, en todo caso, una vez tomada una decisión sobre el lugar a donde ir, será estable.

[30] Gauthier, D., "David Hume Contractarian", en *Moral Dealing* (cit.), pp. 45-76; p. 49.

[31] Platón, *La República*, 338c.

poderosos), ha de ser puesta en acto mediante la instauración de sanciones, tales que no pueden explicarse sino por referencia a un contrato o, al menos, que dotan a la situación final de la apariencia de un contrato. En este caso, una convención estable entre aquellos a quienes beneficia, no lo es referida a la totalidad de los sujetos, y ha de ser disfrazada con el aspecto de un contrato (falaz) de sujeción en beneficio de todos.

Así, como mucho, podemos decir que el contractualismo implica no cualquier contrato, sino un contrato entendido como un pacto hipotético al que podrían asentir *todos*, como agentes racionales perfectamente informados. Mientras que algunos tipos de convencionalismo también habrán de hacer uso del contrato para justificar ciertas instituciones, pero se tratará de un contrato ocasionalmente engañoso y falaz.

2.- Pasamos al segundo criterio de distinción, referido al carácter normativo de ambos tipos de teorías. De un modo un tanto dogmático, diremos que las teorías convencionalistas tienden a ser teorías empíricas sobre cómo se originan y mantienen, de hecho, las instituciones sociales y políticas. A diferencia de ellas, las teorías contractualistas buscan razones que justifiquen racionalmente un modelo determinado de Estado o de legislación. En contractualismo tendría, así, un carácter "transcendental", frente al carácter empírico del convencionalismo. Éste se conforma con extraer conclusiones *a posteriori*, mientras aquél pretende ofrecer principios normativos *a priori*. No en vano el paradigma del contractualismo hipotético ha sido, en nuestro siglo, la teoría de la justicia de Rawls, autocalificada como constructivismo kantiano.

En un sentido similar, Gauthier ofrece una interesante lectura de Kant en "The Unity of Reason: A Subversive Reinterpretation of Kant"[32]. Su conclusión es que, si Kant hubiese buscado las condiciones de posibilidad de la acción como buscó las del conocimiento (es decir, si hubiera reproducido su filosofía teórica en la práctica), sin exceder los límites de la experiencia posible, habría hallado leyes prácticas *a priori* como halló conceptos puros del entendimiento. No nos detendremos aquí en el argumento de Gauthier, simplemente diremos que el procedimiento por el que, según él, cabe extraer estas leyes prácticas *a priori* —especialmente en cuanto se refieren a la interacción— es un procedimiento contractualista. Tal procedimiento partiría de examinar el papel que juega la razón práctica en la relación entre acción y la multiplicidad de los deseos, paralelo al que juega la razón teórica en la

[32] en *Ethics*, nº 96 (1985), pp. 74-88, reimpreso en *Moral Dealing*, cit., pp. 110-126, por donde se citará.

establecida entre el conocimiento y la multiplicidad de la intuición[33].

Según estas lecturas, el contractualismo sería el procedimiento por el que la razón individual se remonta hasta encontrar leyes prácticas (derivadas de las primeras leyes prácticas *a priori*) para la interacción con otros sujetos. La razón juega —en el marco del sujeto transcendental, habría que decir— un papel creativo al unificar los deseos para, de acuerdo con el principio de felicidad, constituir una acción con sentido. De igual modo, en un segundo nivel, la idea de una racionalidad común (expresada procedimentalmente en la "historia" de una negociación entre las racionalidades individuales) unifica la multiplicidad de intereses y proyectos de felicidad individuales para constituir una sociedad que los haga compatibles en la mayor medida posible, de acuerdo a ciertos principios. El contractualismo vendría a ser la filosofía transcendental de la sociedad, aquél ejercicio que desvela las condiciones de posibilidad de cualquier sociedad posible[34]; mostrando, además, la actividad creadora de la razón, al establecer los principios *a priori* sobre los que se asienta esa posibilidad. El contractualismo no sólo se revela como una teoría normativa, sino que muestra el carácter auto-legislador de la razón individual, generalizada mediante el expediente del contrato.

En principio, este enfoque transcendental no implica necesariamente el contractualismo como modelo teórico para descubrir los principios de una sociedad justa. De hecho, el más kantiano de los contractualistas, Rawls, no emplea la idea de contrato en un sentido estricto, como ya hemos reiterado. Sin embargo, el método contractualista es el único que capta adecuadamente la naturaleza procesal y plural de la razón común; de otro modo, el punto de vista

---

[33] Este paralelismo acabaría por asignar a la felicidad un papel muy diferente al que Kant le otorgó, pues "la felicidad, la satisfacción de todos los deseos, es dada como el fin de la acción no, como Kant parece haber supuesto, por necesidad natural, sino como el resultado de la actividad de la voluntad, o razón práctica, al unificar la multiplicidad del deseo para determinar una única acción [...]. Dado que es la razón la que unifica la multiplicidad de los deseos para posibilitar la elección, es ella la que prescribe la realización de aquella acción que maximice la satisfacción del agente. La necesidad de buscar la felicidad depende, así, de la actividad sintética *a priori* de la razón práctica. y ese sería el carácter de la ley práctica basada en la felicidad, cuya posibilidad Kant niega." (Gauthier, D., "The Unity of Reason..." cit., p. 116.

[34] El contractualismo no parece responder a la pregunta empírica de cuál es el origen histórico o social de nuestras reglas políticas y morales, o de qué consenso fáctico —expreso o tácito— extraen las normas e instituciones políticas su legitimidad social (aunque se presente así muchas veces, en la versión denominada "realista" o "histórica" del contrato); más bien parece dirigirse a contestar la pregunta de cómo son posibles en general el Estado, la sociedad, la justicia, la obligación política, el poder y las reglas políticas o morales, dada la naturaleza individual humana para, partiendo de esa misma base de la naturaleza humana, averiguar qué reglas o instituciones son más plausibles, razonables o justas, y de este modo, legitimarlas racionalmente.

de "razón común" adoptado, sería objetable. Por otro lado, es el método que viene exigido por la estructura de la interacción racional, como quedó expuesto, para escapar a sus dilemas. Además, es el único que permite deducir los principios de la justicia a partir de las racionalidades y autonomía individuales, concediendo (a través de las hipótesis de la negociación ideal y el contrato) igual peso a los intereses y proyectos de cada individuo. Así, la hipótesis del contrato resulta el método que mejor capta el carácter plural de la razón común y que más razonablemente puede acercarnos a los principios universales (racionales) de la justicia.

En definitiva, el contractualismo posee (o pretende poseer) un papel normativo *a priori* que el convencionalismo no persigue. No obstante, debemos introducir algún matiz en este enfoque "kantiano", pues muchas teorías políticas contractualistas difícilmente pueden considerarse normativas en un sentido tan estricto. Por ejemplo, gran parte de las teorías políticas medievales elaboradas para legitimar la soberanía mediante el recurso a la hipótesis de un *pactum subjectionis* expreso o tácito, pueden considerarse contractualistas, pese a que no pretenden descubrir principios políticos universales, sino únicamente legitimar instituciones concretas sobre la base de un asentimiento histórico o hipotético. Estas teorías están a medio camino entre el convencionalismo y el contractualismo tal como vamos dibujándolos hasta ahora. Si no produce mucha confusión, podemos identificarlas como teorías *pactistas*[35], para diferenciarlas igualmente del contractualismo propiamente dicho y del convencionalismo, de las que están igualmente alejadas.

3.- Como tercer criterio de distinción aludíamos al racionalismo. Consideramos que es este el criterio más útil para diferenciar las teorías convencionalistas y las contractualistas y, sin embargo, es difícil de explicar sin caer en una simplificación que sería injusta, especialmente con el convencionalismo. El criterio puede enunciarse diciendo que, por regla general, las convenciones no requieren una justificación racional fuerte. Como acontece en el traído y llevado ejemplo de las normas de circulación, su contenido es perfectamente arbitrario o voluntario. Por el contrario, los contratos (puesto que imponen cargas) deben obedecer a razones; necesitan estar racionalmente fundados.

Se dirá que las convenciones también tienen una base racional, pues la necesidad de coordinación lo es. Sin embargo, el contenido mismo de la convención es puramente arbitrario. Una sociedad que no lograra ponerse de

[35] O, también "contractualismo político", frente al "contractualismo filosófico" que es propiamente el que consideramos paradigmático.

acuerdo sobre si conducir por la izquierda o por la derecha manifestaría una peligrosa irracionalidad, pero una vez que se ha seleccionado un principio de coordinación, este es inmune a la crítica racional. Se puede decir que depende de las libres voluntades de los agentes y, desde ese punto de vista, no hay base racional para la crítica.

No ocurre así con los contratos. Estos suponen una causa al menos razonable, y su contenido no es arbitrario. Es más, incluso tras la coincidencia de voluntades entre las partes, puede averiguarse —como se hace usualmente en derecho— si el contrato mismo es equitativo, esto es, razonable y, en caso contrario, cabe denunciarlo (incluso contra la primera voluntad de los contratantes cuyos intereses lesione). Lo que queremos decir es que el contrato, por su naturaleza, no busca una simple coordinación, sino algo más (la *cooperación*). Busca realizar un intercambio que, pese a imponer cargas en todas o algunas de las partes, sea aceptable para ellas. Y para lograr esa aceptabilidad necesita apelar no sólo a su voluntad o su arbitrio, sino a su racionalidad (esto es, a su interés).

Esta diferencia entre convención y contrato se traduce fácilmente a las respectivas teorías políticas o filosóficas que los emplean como paradigmas del origen y fundamento de las instituciones políticas o morales. El convencionalismo se distingue por apelar a las voluntades de las partes, mientras el contractualismo apela a su racionalidad (su verdadero interés). Ello puede ser consecuencia de que el convencionalista no escapa al escepticismo inicial que le conduce a buscar una justificación pactada a ciertas instituciones, mientras que el contractualista, pese al escepticismo inicial, cree hallar, en su método, a través de la racionalidad individual, un camino seguro para retornar a la confianza en la razón. A su vez, estos distintos enfoques conducen al convencionalista a prestar más atención a las pasiones humanas, y así es común que el convencionalismo se asocie a la convicción de que la voluntad de poder o dominio, o la concupiscencia de cualquier índole, gobierna irremediablemente a los hombres. Ocasionalmente, el contractualismo concede también importancia a estas pasiones, pero siempre acaba reconociendo un lugar para la restricción racional: los hombre aparecen como seres racionales capaces de domeñar sus tendencias instintivas por su propio interés. Las teorías estrictamente contractualistas acaban siendo teorías sobre la naturaleza de la razón, como ya sugería en el "enfoque transcendental" expuesto arriba.

Otra consecuencia de este racionalismo propio de las teorías del contrato es que estas teorías tratan de explicar la generalidad o universalismo de las reglas morales o de los principios de justicia, suponiendo que tales principios

no son meras convenciones, sino verdades necesarias o permanentes[36], susceptibles de ser descubiertas mediante el uso de la razón. Por el contrario, es más propio del convencionalismo —partiendo de que los principios son pactados (y, por ende, contingentes)—, centrarse en explicar y justificar el relativismo de los valores morales y principios políticos.

Estos criterios ayudarán, sin duda, a comenzar a distinguir qué teorías pueden considerarse convencionalistas y cuáles contractualistas. No obstante, antes de ofrecer una versión más acabada y precisa de dicha distinción, aludiremos a otras diferencias posibles, que complementarán las hasta aquí expuestas.

d) Distinciones complementarias.-

Muchos autores, y en especial los pertenecientes a la tradición del contrato social, median en este problema de la precisión entre convencionalismo y el contractualismo, para evitar la frecuente y dañina confusión entre ellos. Arriba veíamos como Hume, por ejemplo, dedica un opúsculo[37] al contrato social, donde hace algunas clasificaciones interesantes. Ahora veremos qué

---

[36] Hampton aclara este punto para nosotros al comentar la ética de Hobbes: "El mismo hecho de que Hobbes escribiera el *Leviathan* evidencia que, al final, rechazó la idea de que las verdades morales fuesen sólo convencionales. El *Leviathan* presenta, después de todo, un elaborado argumento en defensa de la segunda ley de la naturaleza, que ordena a los hombres instituir un soberano para alcanzar la paz. La verdad de esta ley no es convencional; es verdadera porque describe correctamente la conexión causal entre la paz y las acciones que dan lugar a la institución del soberano (...). Así, si no hubiera verdades morales fuera de aquellas convencionalmente establecidas, el proyecto de Hobbes en el *Leviathan* sería imposible. El considerable caos moral que tanto le preocupa sería irremediable, porque el productor último de la paz, el soberano, no podría ser instituido. La más profunda fe de Hobbes fue que podía mostrar una *conexión causal necesaria* entre, al menos un proyecto cooperativo (la institución del soberano) y la paz." (*Hobbes and the Social Contract Tradition,* cit., p. 49). Sin embargo, debemos matizar que Hampton desgaja la ética de Hobbes de su filosofía política, con lo que no la considera, como sí intentaremos nosotros más abajo, una moral contractual. Según nuestra hipótesis, las leyes de la naturaleza, que Hampton denomina verdades morales, serían más bien mandatos de la razón, aún sin contenido moral propiamente dicho. Pero esto no modifica el sentido del párrafo que hemos reproducido: la idea de una conexión causal necesaria entre ciertos fines (individuales o colectivos) y ciertos medios (necesariamente colectivos y cooperativos) es la dovela clave del contractualismo.

[37] Se trata de *Of the Original Contract*, incluido en *Essays and Treatises on Miscellaneous Subjects*, Londres, 1777, pt II, nº XII.

interés puede tener la visión de Gauthier sobre este problema. Su opinión en este punto está diseminada en varios artículos sobre el contractualismo histórico[38]. Quizá donde más claramente la establece es, casi de pasada, al comienzo de "David Hume, Contractarian", donde dice sencillamente: "El contractualismo es una especie de convencionalismo normativo"[39].

Según este punto de vista, el convencionalismo es un género muy amplio, que abarcaría cualquier teoría que explicase la acción por referencia al interés común o la utilidad pública que cada uno reconoce y que le lleva a concurrir con los demás en un plan general de acciones. Dentro de este género, el contractualismo es una especie; pero el utilitarismo también puede entenderse incluido en el mismo[40].

Lo que distingue a esta especie de convencionalismo normativo —que da lugar a "convenciones contractuales"— es que en él están presentes mecanismos tales como la negociación y el pacto (innecesarios en otros tipos de convención); entendiendo "por *negociación* un acuerdo en el que cada persona entra por su propio interés, que resulta en la selección de una convención", y "por *pacto* también un acuerdo en el que cada persona entra por su propio interés, que asegura, con o sin coacción, la adhesión mutua a la convención"[41]. La negociación implica la resolución de preferencias opuestas, que se efectúa mediante el mutuo reconocimiento de las mismas; el pacto implica una obligación racional para el obligado (favorable a sus intereses) incluso aunque en el momento de realizarse suponga el abandono de la estrategia maximizadora y deba, por ello, asegurarse mediante coacción. Así, Gauthier concluye que una convención es un contrato si y sólo si es seleccionada entre las alternativas posibles mediante un reconocimiento de intereses o prescribe su propio cumplimiento sobre la base de una obligación auto-interesada[42].

En cierta manera, la perspectiva de Gauthier reproduce algunos de los criterios que comentábamos arriba. En esa medida, es una fórmula aceptable para distinguir convenciones de contratos. Sin embargo es, como puede verse,

---

[38] Especialmente en "Thomas Hobbes: Moral Theorist", cit, "David Hume, Contractarian", cit. y "The Social Contract as Ideology", *Philosophy and Public Affairs*, 6 (invierno 1977), pp. 130-164, reimpreso en *Moral Dealing*, cit., pp. 325-354.

[39] p. 45.

[40] Cfr. Gauthier, D., "David Hume, Contractarian", cit., p. 45.

[41] Gauthier, D., "David Hume, Contractarian", cit., p. 50.

[42] Cfr. Gauthier, *ibidem*.

una distinción excesivamente formal, que no refleja directamente los dos criterios últimos (carácter normativo y racional del contractualismo) que nosotros hemos destacado. Al hablar de convención normativa, Gauthier sólo alude a la distinción entre convenciones o contratos analíticos y legitimadores, según sus pretensiones[43], pero no añade ningún rasgo sustantivo para diferenciarlas.

Nosotros debemos insistir en esos caracteres que, aunque no dan lugar a grandes diferencias en la forma, sí comportan una profunda distancia en el alcance de ambos paradigmas, hasta el punto que consideramos dudoso que puedan tratarse como género y especie.

La distancia entre convención y contrato tal vez aparezca más clara por referencia a su aplicación, es decir, al uso filosófico para el que se disponen. Hemos distinguido al menos tres papeles filosófico-políticos fundamentales que pueden desempeñar las teorías contractualistas y convencionalistas en general:

1.- Un *papel justificador*; consistente en ofrecer al individuo un medio para justificar racionalmente (ante sí o ante otros) su comportamiento moral y político, su participación en la creación y conservación de ciertas instituciones o, incluso, el compromiso para adoptar cierto "carácter moral"; y en ofrecer a las instituciones un medio igualmente racional para reclamar la obediencia (obligación política) de todos los agentes racionales[44].

2.- Un *papel epistemológico*, consistente en mostrar cómo únicamente es posible conocer qué principio moral es correcto, o qué principio político justo, o qué gobierno legítimo; pero sin derivar conclusiones normativas directamente.

3.- Un *papel descriptivo*, o genético, consistente en elaborar hipótesis sobre el origen probable (histórico o lógico) de los Estados o las reglas morales.

Cada uno de estos posibles papeles tiene distintas lecturas según el paradigma elegido sea convencionalista o contractualista. Ambos paradigmas

---

[43] Ya aludida arriba, en los puntos *1.* y *2.*a) de este capítulo.

[44] Sobre el escurridizo concepto de justificación, seguimos, principalmente, a Kurt Baier; cfr. su "Justification in Ethics", en Pennock, J.R., y Chapman, J.W. (eds.), *Justification* (*Nomos*, vol. XXVIII), Nueva York, New York University Press, 1986, pp. 3-27.

pueden cumplir las tres funciones —con resultados dispares— pero el contractualismo es más adecuado a la función epistemológica y justificadora, mientras el convencionalismo rinde mejores resultados en una función descriptiva. No obstante, no hay que descartar su virtualidad como medio de justificación, mediante su apelación a las convenciones efectivamente mantenidas (por ejemplo, a la "moral social") y la coacción a ellas asociada. Sin embargo, como fundamento epistemológico supondría una teoría convencionalista de la verdad, o simplemente una postura escéptica.

Los tres papeles o niveles de aplicación de las teorías filosófico-políticas y éticas nos permiten ahora una distinción algo más precisa, basada en la referencia a cómo cumplen convencionalismo y contractualismo, respectivamente, los papeles motivador (asociado a la justificación), epistemológico y descriptivo-genético.

e) Motivación, conocimiento y explicación.-

Bajo este epígrafe vamos a intentar esquematizar las diferencias que supone la aplicación de los paradigmas convencionalista y contractualista, respectivamente, a estos tres problemas centrales de la filosofía práctica identificados en el epígrafe anterior. Para ello, sólo enunciaremos brevemente la respuesta de cada paradigma, pues consideramos que su simple exposición, y la comparación con el paradigma alternativo, son comentario y explicación suficiente:

1.- El paradigma convencionalista:

Motivación convencional: Como agente, debo hacer la acción *A* en una situación *S* porque la mayoría así lo hace, con la opinión de que es correcto, y la convicción de que es un comportamiento común y esperado en la situación *S*.

Convencionalismo epistemológico: Posibilidad *a)*, (escepticismo) no hay una verdad en materia práctica; *aceptamos* como correcto lo que ha sido pactado de hecho, o lo que se observa mayoritariamente como si hubiese

sido pactado. Posibilidad *b)*, (convencionalismo epistemológico) la verdad en materia práctica es convencional: lo correcto *es*, o *equivale a*, lo convenido.

Génesis convencional: La moral o el Estado *es fruto de una convención*. Esta convención ha sido producida por un pacto real entre los hombres (tal vez incluso histórico), o mediante un proceso espontáneo de coordinación.

2.- El paradigma contractualista:

Motivación contractualista: Como agente, debo realizar la acción correcta. La acción correcta se define, bien como aquella cuya máxima no podría haber sido rechazada por ninguno de los demás como base para un acuerdo unánime[45]; bien —en la versión de Rawls— como aquella que cumpla con una regla susceptible de haber sido pactada unánimemente por agentes racionales en condiciones ideales e imparciales de elección.

Contractualismo epistemológico : Hay una verdad en materia práctica, racionalmente cognoscible. Alcanzamos ese conocimiento mediante la reconstrucción racional del acuerdo originario de la moralidad y el Estado. Esta reconstrucción es transcendental en cuanto se refiere a las condiciones de posibilidad de cualquier regla moral o política correcta.

Génesis contractual: El contractualismo no se preocupa de la génesis

[45] Este enunciado está basado en la idea de contractualismo de T.M. Scanlon en "Contractualism and Utilitarianism", en Sen, A. y Williams, B. (eds.), *Utilitarianism and Beyond*, Cambridge, Cambridge University Press, 1982, pp. 103-128.

histórica de las estructuras jurídicas, política o morales. Se centra en su legitimación o justificación racional. Para ello, realiza una reconstrucción racional, no histórica.

f) Conclusión: nuestra visión de la diferencia.-

Como final de este punto queremos ofrecer una breve conclusión sobre los datos y reflexiones que hemos aportado. La siguiente tabla, que resalta algunas de las coincidencias y diferencias entre los paradigmas que venimos comparando, puede orientará nuestra conclusión:

contractualismo	convencionalismo
esceptisimo inicial	escepticismo inicial
confianza en la razón	desconfianza de la razón
racionalismo	voluntarismo/decisionismo
universalismo	relativismo
reconstrucción racional	reconstrucción histórica

La primera fila de la tabla muestra el común punto de partida de ambos enfoques, que tematizamos en el epígrafe *b)*. La discrepancia comienza a mostrarse a partir de la segunda fila. De todo cuanto hemos dicho se deriva que la gran diferencia entre estos paradigmas reside en que el contractualismo representa un proyecto filosófico basado en la confianza de la razón (la razón es un medio válido para escapar a sus propias contradicciones), mientras el convencionalismo no supera el inicial escepticismo, y permanece anclado en la desconfianza hacia la razón. El convencionalismo es escéptico de principio a fin: la común idea inicial de que el hombre está guiado únicamente por sus intereses se radicaliza entre los convencionalistas, quienes creen que no podemos escapar a esta determinación, por lo que la única instancia normativa podrá ser una contingente convención de hecho mantenida coactivamente. Por el contrario, el contractualista encuentra en el auto-interés una expresión de la racionalidad individual que, si bien define al hombre y lo conduce a los conocidos dilemas de la racionalidad colectiva, también le indica una salida

racional al conflicto natural. A través de esa salida, el complejo normativo que se crea es capaz de ofrecer una justificación de la acción moral enteramente independiente de las sanciones (aunque, como mecanismo político, la ley positiva y su sistema de sanciones pueden ser también legitimados contractualmente).

El contractualismo supone, así, el establecimiento de un orden normativo basado en la idea de una reconstrucción racional, que adopta la forma extendida de una negociación y un pacto entre agentes racionales independientes. Mientras, el convencionalismo supone el establecimiento de un orden normativo de raíz puramente decisionista o voluntarista —pues la racionalidad es siempre únicamente individual (egoísta, a-social)— pero, como señala Gauthier, tal orden convencional se ha de venir abajo cuando, incapaz de generar una auténtica "razón común", se revela su radical impostura[46].

Encontramos, pues, la diferencia básica entre ambos paradigmas en el peso relativo que conceden a la racionalidad o a la mera voluntad individuales. El contractualismo encuentra una virtualidad en la racionalidad individual: la posibilidad de, mediante un proceso constructivo, engendrar (o, mejor, reconocer) la posibilidad de una razón común, una instancia normativa intersubjetiva. El convencionalista no admite esta posibilidad y, por tanto, la normatividad intersubjetiva es, para él, una mera ficción de "razón común", una "buena estrategia" al servicio de cada individuo, pero basada en una decisión contingente. Para el contractualista *hay* una conexión causal necesaria entre autointerés y principios morales (o políticos); para el convencionalista no.

Como consecuencia de esto, el contractualista cree estar construyendo, con su teoría, un argumento de validez universal: una justificación racional de ciertos principios constante y generalmente válidos. Si sus premisas son correctas, sus conclusiones substantivas son universales. Muy otro es el alcance del convencionalismo: si sus premisas son correctas, el universalismo es una quimera. Todo lo que podríamos ofrecer son justificaciones parciales y de

[46] Gauthier, D., *MA*, pp. 316-317: "Podría ser que las características de la naturaleza humana que intervienen en la fundación de estas restricciones [morales] sirvieran también para socavar las condiciones bajo las cuales sería racional cumplirlas. Podría ser que necesitásemos restricciones morales dada nuestra a-socialidad y nuestro desinterés por los demás (*non-tuism*), pero que, precisamente por ser asociales y desinteresados en los demás, diésemos lugar a circunstancias en las cuales la pretendida imparcialidad de estas restricciones fuera una impostura. Y si la impostura es reconocida, las restricciones dejan de ser efectivas."

validez relativa[47].

El convencionalismo supondrá inevitablemente que cualquier convención que no sea meramente coordinativa (como el acuerdo entre los dos remeros) está condenada a ser impuesta mediante coacción o engaño (aunque sea auto-engaño). En un contexto convencional, se cumplirían las previsiones de Trasímaco y Glaucón: ante la mínima oportunidad de cometer injusticia, como la que tiene el pastor lidio al portar el anillo de Giges[48], no habría razón alguna para comportarse justamente. Pero el contractualismo, al dirigirse a la razón de cada individuo, con la autoridad de estar fundado en ella misma, pretende haber establecido un fundamento normativo firme de carácter interno. La coacción es una estrategia necesaria sólo porque nuestra racionalidad no es perfecta: en muchas ocasiones los cálculos auto-interesados a corto plazo, o las pasiones, nos apartan de nuestro interés considerado. La coacción es un remedio contra la flaqueza de la voluntad, no una necesidad impuesta por la contingencia inevitable de nuestras convenciones normativas.

Desde nuestro punto de vista, lo que acabamos de exponer es el punto clave para la comprensión de nuestra hipótesis de que, si bien es posible hablar de un contractualismo moral —aunque veremos con qué limitaciones y en qué escasos ejemplos— es imposible hacerlo, rigurosamente, de un convencionalismo moral. Una moral convencional es, sencillamente, algo distinto de una moral, que sólo por analogía puede recibir tal nombre.

Por eso, sostendremos que las versiones de la teoría hobbesiana que consiguen mantener su mismo esquema y estructura argumental en ausencia de un soberano coactivo[49], no sólo ponen de manifiesto claramente el carácter contractualista de la misma, sino que establecen, además, las bases para su posible lectura como contractualismo moral[50].

---

[47] Estos comentarios sugieren una diferencia adicional, pues el contractualismo aparece como una teoría capaz de ofrecer un contenido substantivo de la justicia y la moral (aunque sea a través de la selección de una regla o principio general *cuasi*-formal). El convencionalismo es, sin embargo, puro formalismo: será "justa", literalmente, *cualquier regla o institución* fruto de una convención. No hay límite o norma para su contenido.

[48] Cfr. Platón, *La República*, 359*d*.

[49] El paradigma de este tipo de lectura es G. Kavka. Cfr. su obra *Hobbesian Moral and Political Theory*, Princeton, Princeton Universtiy Press, 1986; esp. caps. 4, 8 y 9.

[50] Que sería, anticipamos, una lectura subversiva y un tanto libre.

El mayor reto del contractualismo, y su "prueba de fuego", consiste en responder satisfactoriamente a la famosa objeción del Necio (*Foole*)[51]. Al hacerlo, está demostrando que el contrato es algo más que una conveniencia mantenida por la fuerza; es un acto racional, la necesidad de cuya observancia está necesariamente implicada en su misma producción y resulta evidente para cualquiera que razone rectamente. El contractualista no amenaza al Necio, ni lo seduce para que se adhiera a su "partido"; persuade a su razón, mostrando la necesidad a que está sometida su naturaleza auto-interesada de erigir una "Razón Común" y atenerse a ella en las controversias prácticas[52]. La objeción del Necio debería diluirse así, según el contractualismo.

Desde el punto de vista convencional, la objeción es, por el contrario, completamente pertinente, acertada e insuperable. Por eso decíamos que el éxito en responder a ella (o, al menos, el intento fundado de hacerlo) es la prueba de fuego del contractualismo. El contractualismo se distingue por persuadir al Necio con argumentos; el convencionalismo sólo le convencerá si puede apelar a la coacción o al engaño.

Como conclusión de nuestra reflexión, hemos de reconocer la aparentemente corta distancia que separa a dos paradigmas filosóficos radicalmente diferentes. Las premisas comunes, el procedimiento argumental y el lenguaje típicamente empleado son coincidentes. No es ahí donde encontraremos bases suficientes para la necesaria distinción. Habremos de fijarnos en el alcance u objetivo de las teorías: las contractualistas se plantean como un intento de descubrir qué es individualmente racional hacer, de acuerdo a qué principios o en qué marco institucional es racional actuar (lo que implica que es racional contribuir a que existan), teniendo en cuenta la naturaleza humana. Las convencionalistas se contentan comúnmente con explicar el relativismo de los valores y las leyes humanas, así como justificar el poder de hecho, mediante la explicitación de los acuerdos o asentimientos expresos o tácitos que lo mantienen.

La venerable teoría del contrato social aparece, así, propiamente entendida, como algo muy diferente de un mero "convencionalismo normativo"; como una estrategia de la razón encaminada a justificar ante el individuo (único soberano) la obligación política y moral; como *ratio cognoscendi* de la morali-

[51] Cfr. Hobbes, T., *Leviathan*, cap. XV.

[52] Cfr. Hobbes, T., *Leviathan*, cap. V.

dad. Ambos objetivos, justificador y epistemológico, se logran mediante una reconstrucción que parte de un compromiso radical *en contra* de lo que reconstruye, a fin de asegurar la pureza del proceso heurístico que debe conducir, desde sus elementos independientes, necesariamente, hasta la estructura que ha de ser justificada. Así, la justificación del Estado y la obligación política se inicia con una presunción en contra de la sociabilidad y en contra del interés en los demás que pudiera servir de base a cierta obediencia. De igual modo, la justificación de principios morales requerirá una presunción en contra de la moralidad, una radical renuncia a la presencia de premisas morales entre los elementos del contrato. Y ahí es donde está la clave de nuestra lectura de la historia del contractualismo, pues mostraremos que, excepto Hobbes (y con muchas reservas) ningún contractualista adopta estas radicales presunciones. Por eso, no podremos hablar en rigor de contractualismo moral hasta nuestro siglo.

En el punto siguiente asistiremos a varios ejemplos de cómo el convencionalismo político (e incluso moral[53]) es moneda de uso corriente en la historia de la filosofía política y moral. Lo mismo que el contractualismo político. Sin embargo, veremos que el contractualismo moral apenas puede ser sospechado en algunos pasajes de Rousseau y, casi a la fuerza, impuesto a los textos de Hobbes.

[53] El hecho de que, como decíamos arriba, una "moral convencional" sea casi una contradicción en los términos, no impide que existan argumentos morales convencionalistas: especialmente sobre el origen y carácter de la moralidad. En realidad son argumentos escépticos y, en última instancia, auto-contradictorios, pero sus ejemplos no escasean en la historia de la filosofía.

3. Contractualismo político y convencionalismo moral

a) Glaucón, ¿el primer contractualista?.-

Un contrato o una convención son extrañas respuestas a una extraña pregunta. En efecto, la formulación de la pregunta por la justicia en *La República*[54] de Platón debió parecer sorprendente. Para cualquier ateniense, lo justo es lo que es conforme a la ley. La auténtica pregunta es más bien esta otra: ¿es más virtuoso el justo o el injusto, el que cumple siempre la ley o aquél que tiene grandeza suficiente para pasar por encima de ella[55]? Y esta es la pregunta que, inconscientemente, responden Polemarco, Trasímaco y, luego, de alguna forma, Glaucón y Adimanto. Esta confusión es más que una anécdota; tiene un significado extraordinario. En las respuestas de los sofistas y en la insistencia de Sócrates se va abriendo camino la necesidad de encarar la pregunta efectivamente planteada, ¿qué es la justicia?, en el sentido de ¿qué ley es justa?, pues hay al menos dos tipos de leyes, la incorporada en el orden natural (*physis*) y la ley civil (*nomos*). La justicia se refiere usualmente al *nomos*, a una norma convencional, que cambia de año en año, de ciudad en ciudad, ¿qué significado, que no sea trivial o relativo, puede tener, entonces, la justicia?, y ¿puede conformarse el filósofo con ese significado? Los sofistas ofrecen sin rebozo explicaciones de este tipo: la justicia es la voluntad del más fuerte; la justicia es la voluntad de los poderosos, etc.

Tales respuestas muestran, en realidad —de acuerdo, por otro lado, con el plan e intención de Platón— el estado de corrupción del lenguaje sobre la

---

[54] Todos los pasajes de *La República* que citaremos y aludiremos en lo que sigue, son transcritos según la versión castellana de la edición bilingüe de José Manuel Pabón y Manuel Fernández Galiano, Madrid, Instituto de Estudios Políticos, 1969, en tres volúmenes.

[55] A pesar de las revoluciones legislativas de Dracón y Solón, los atenienses debían concebir las leyes escritas como un mal necesario. Hay que recordar que su paradigma de virtud y excelencia (identificado con la *areté* homérica, y herencia suya) tiene su ámbito en un mundo sin leyes escritas, donde la admiración y la calificación de *agathós* (bueno) se dirige a quien muestra capacidad de sobresalir, de escapar a la norma común.

justicia en los tiempos de la sofística. En el diálogo platónico, la intervención de Glaucón tiene ese mismo propósito, salvo que tal vez representa la versión más refinada del escepticismo. En cualquier caso, es expuesta para ser refutada por Sócrates[56]. Nuestro análisis debe, para valorar su significado como primer antecedente histórico del contractualismo o del convencionalismo, hacer abstracción de ese contexto. Leeremos la teoría de Glaucón sobre el origen de la justicia como el intento más serio de establecer una alternativa sofística (y, por tanto, ilustrada y crítica) al decadente paradigma mítico y heroico tradicional.

En primer lugar, hay que constatar que, efectivamente, el personaje de Glaucón desarrolla sistemáticamente, por primera vez, una teoría aparentemente contractual del origen de la justicia[57]. Aunque su lenguaje está "contaminado" por la reciente discusión entre Sócrates y Trasímaco[58], podemos distinguir en su teoría la figura de un "estado natural" en que cada cual actúa como le place, siguiendo —como escribe un poco más abajo— "...el interés propio, finalidad que todo ser está dispuesto por naturaleza a perseguir como un bien, aunque la ley desvíe por fuerza esta tendencia..."[59]. También reconocemos en Glaucón el fundamento racional del contrato, pues, al probar los hombres las desventajosas consecuencias de su libre actividad, cada cual reconoció que

> "lo mejor era establecer mutuos convenios con el fin de no cometer ni padecer injusticias. Y de ahí en adelante empezaron a dictar leyes y concertar tratados recíprocos, y llamaron legal y

---

[56] Y, por otro lado, el propio Glaucón insiste en que está transmitiendo una idea impía, alejada de sus propias convicciones. Así, por ejemplo, en 361*e* leemos: "no creas, Sócrates, que hablo por boca mía, sino en nombre de quienes prefieren la injusticia a la justicia...". Estas prevenciones no desvirtúan la potencia del argumento (más bien lo realzan, ya que hay que suponer que se hacen necesarias debido a su propio poder de convicción o verosimilitud).

[57] Cfr. Platón, *La República*, 358*e* y 359*a*.

[58] Y así habla de cometer y sufrir injusticias antes de que existan comunidad civil y leyes.

[59] *La República*, 559*c*. Gauthier ha reconocido en Glaucón un estado de naturaleza, aunque matizando, también, el problema de la "contaminación" terminológica a que nos referíamos. Gauthier constata que Glaucón "habla de personas que sufren injusticia antes de aceptar las restricciones que constituyen la justicia, cuando, más correctamente, debería decir que, aunque las personas naturalmente hacen lo que vendrá a ser considerado injusto, originalmente están situadas fuera de toda restricción moral (...). Si la moralidad es materia de un acuerdo, entonces antes del acuerdo nada es justo ni injusto" (*MA*, p. 309-310).

> justo a lo que la ley prescribe. He aquí expuesta la génesis y esencia de la justicia." (359*a*)

Si el Glaucón platónico transmite fielmente las teorías predominantes entre algunos de los sofistas, se puede decir que, efectivamente, la tradición del contrato social (en sentido amplio) se originó entre ellos. Sin embargo, las características del breve esbozo (apenas unas líneas) de Glaucón no nos autorizan a afirmar —como con cierta ligereza hace Gauthier— que el "*contractualismo moral* parece haberse originado entre los sofistas"[60]. Entre las razones por las que el esbozo de Glaucón no puede considerarse propiamente contractualista (y menos contractualista moral) destaca su propia afirmación de que está reflejando la opinión de los que creen que es mejor la injusticia. Por otro lado, el texto que reproducimos puede ser engañoso: afirma que la justicia es esencialmente contractual; pero hay que reparar en que, en este caso, el contrato, por su naturaleza esencialmente instrumental (en relación con los fines individuales), no remonta ese mero uso instrumental y no da lugar a una verdadera justicia, sino sólo a la ficción de ciertas convenciones capaces de reprimir el impulso que llevaría a cada agente a realizar su propio bien, cifrado en cometer injusticia y no ser castigado por ello.

Sin la sutileza aristotélica, Glaucón no diferencia lo justo legal y lo justo natural y, de este modo, puede parecer que habla de la justicia esencial, cuando sólo habla de convenciones legales. Sobre lo único que realmente acuerdan los individuos glauconianos es sobre un nombre: "...*llamaron* legal y justo a lo que la ley prescribe".

Lo que describe Glaucón no es, por tanto, la esencia de la justicia, sino el origen de lo que los hombres —tal vez equivocadamente, tal vez contra la naturaleza— han acordado llamar justo. Pero todo esto caracteriza un convencionalismo jurídico y político (y, si se quiere, moral), como refrendan las palabras de Adimanto, un poco más abajo:

> "El mundo entero repite a coro que la templanza y la justicia son buenas, es cierto, pero difíciles de practicar y penosas, y en cambio, la licencia e injusticia son agradables, es fácil conseguirlas, y si son tenidas por vergonzosas es *únicamente porque así lo imponen la opinión general y las convenciones.*"[61]

---

[60] *MA*, pp. 9-10, subrayado mío.

[61] *La República*, 364*a*, subrayado mío.

Así, para el sofista, la única regla inquebrantable es la ley natural, aquella que nadie va a incumplir ni siquiera cuando está oculto de las miradas de los demás: la que ordena conservar la vida y promover el interés propio. Todas las demás restricciones y reglas sociales, son tenidas por meros instrumentos al servicio de ese fin supremo[62]. En este sentido, el hombre naturalmente libre se halla constreñido por reglas que —en la medida en que no sólo no reflejan, sino que se oponen a la única ley natural de la libertad— tienen un fundamento puramente decisionista y convencional[63]. El hombre natural de los sofistas no encuentra ningún nexo racional entre el cumplimiento de las reglas sociales y su interés propio, como no sea la necesidad de aparecer como justo ante sus semejantes para así conseguir realizar mejor sus intereses. Por otro lado, el contenido de esas reglas es percibido como completamente arbitrario.

El convencionalismo legal vino a ser un lugar común en la época de Platón, hasta el punto de influir en algunas de sus obras políticas[64]. En todo caso, fue un episodio discontinuo cuya potencia como novedad filosófica entre los sofistas dejó paso a cierta indiferencia: el convencionalismo legal y político era, ya para Aristóteles y, desde luego, para la mayoría de las escuelas helenísticas, una más de las desagradables características del mundo, que había que sobrellevar sin conflictos, pero que carecía de influencia directa sobre la felicidad individual, dependiente de las leyes de la naturaleza.

En definitiva, la sofística creó el lenguaje del convencionalismo legal, político y moral, de modo que representa el primer hito significativo en el

---

[62] Gauthier ha subrayado que la comprensión sofística del individuo que se refleja en esta visión de la ley y la justicia representó una innovación filosófica de primer orden. Así escribe, "de hecho los sofistas captaron, por primera vez en el pensamiento humano, el punto de vista de una persona situada fuera de la vida social, no en sus capacidades, no por ser capaz de vivir sin sociedad, sino en sus motivaciones, en ser capaz de ver la sociedad como un instrumento para fines que no requieren una vida social para ser formulados." (*MA*, p. 312). En el mismo sentido, Cfr. McIntyre, A., *A Short History of Ethics*, Nueva York, 1966, p. 18.

[63] El propio Gauthier —a pesar de su "desliz" inicial— reconoce, en la parte final de *MA*, que "Glaucón retrata la justicia como *una imposición social, convencional* sobre la naturaleza" (p. 309, subrayado mío).

[64] Esa es la opinión de G. Klosko (Cfr. su *The Development of Plato's Political Theory*, Nueva York, Methuen, 1986, p. 231), quien —a pesar de la explícita afirmación de Platón (*Las Leyes*, 889*e*) sobre el origen natural de las normas justas— sostiene que la "carga convencionalista" se acentúa en esta obra y en *El Político*.

camino hacia el contractualismo moral, pero el desarrollo filosófico de la idea de un contrato hipotético como instancia de legitimación política habría de esperar algunos siglos.

b) *Communitas* vs. *Societas*: la disputa medieval.-

La idea de pacto o convención está omnipresente en el pensamiento político medieval. En primer lugar, desde Roma se generaliza la idea de que el fundamento de obligación del derecho de gentes reside en su aceptación de hecho (o tácito consentimiento de los pueblos). Por otro lado, el pacto o consentimiento unánime como fundamento de la obediencia al monarca es una tradición heredada no sólo del convencionalismo griego y helenístico, sino de una tradición mucho más influyente en la Edad Media: la historia del pueblo de Israel[65].

El pensamiento político medieval —cuya complejidad nos vemos obligados a ofender y ocultar con estos breves comentarios— ofrece una variedad riquísima de pactos, consentimientos, contratos, acuerdos, concesiones y sometimientos, todos los cuales vienen a aliviar la tensa relación entre un poder religioso establecido sobre una comunidad cristiana universal, un poder real o imperial que intenta establecerse sobre unos territorios unidos por sus intereses políticos, económicos y estratégicos y los pequeños poderes feudales que luchan por mantener su dominio sobre pequeñas zonas, ofreciendo a cambio paz y seguridad. Cada una de estas relaciones de poder se racionalizan mediante la idea de un intercambio justo, de un pacto explícito o tácito. Ahora bien, en lo que se refiere al pueblo, éste aparece como "comunidad cristiana" en relación al poder religioso, y como "asociación o conjunto de súbditos" en relación al poder civil. Cuando, en la baja Edad Media, la realidad del Estado (con el Rey a la cabeza) se va abriendo paso, surge una nueva forma de relación política que demanda una re-formulación de la doble consideración del pueblo.

[65] Una (extensa) relación de pasajes de la Sagrada Escritura en que se alude a diversos tipos de contrato (entre reyes y pueblo, entre Dios y los reyes, etc.) como justificación de la autoridad real y de otros variados órdenes normativos, puede verse en el capítulo XI del *De cive*, donde son aducidos en favor del contractualismo moderno de Hobbes.

Dos paradigmas, el comunitario y el asociacionista, pugnan por adueñarse del sentido de la nueva unidad política. ambos tienen —aparte de sus respectivas connotaciones religiosas o laicas— un origen jurídico. La *communitas* o *universitas* es una figura de origen germánico, que representa los llamados "bienes en mano común". Se define como una *persona ficta* indivisible, de origen que podemos calificar como "natural" (legal o consuetudinario) porque en él no interviene la voluntad de las partes. La *communitas* nace así de modo natural en la vida jurídica y se mantiene con una personalidad ajena (y superior) a los individuos que eventualmente formen parte de ella. La *communitas* se adapta, por tanto, a la concepción tradicional de la Iglesia, como cuerpo místico, instituido por Dios (no por asociación de los fieles), del que todos los cristianos son parte por la gracia, sin que su eventual decisión de unirse o separarse de él, tomada en sí misma, sea nunca requisito suficiente para la efectiva pertenencia.

Por su parte, la *societas* es de origen romano. Se trata de una asociación nacida de una relación privada, convencional. a veces se define simplemente como el nombre colectivo de los individuos que la componen. Aunque la sociedad posea bienes naturalmente indivisibles, cada miembro se considera propietario de una "parte" de los mismos. Cada individuo está facultado, además, para separarse de la sociedad reclamando su parte en la misma.

Evidentemente, un pensamiento heredero —a través de la recepción escolástica— del racionalismo aristotélico y estoico, y comprometido con la legitimación de las crecientes demandas de independencia de los reyes y príncipes nacionales respecto al papado sólo podía elegir la asociación romana como paradigma de la unión política estatal. Y así, desde la baja Edad Media, el Estado se concibe como una asociación privada de individuos titulares de ciertos derechos y capacidades, entre ellas, la de obligarse entre sí. Hacia finales de la Edad Media y comienzos de la modernidad, la idea de que el Estado se puede explicar como una asociación de hombres iguales está extendida como axioma del Derecho Natural y de Gentes.

Tales ideas, así como los conceptos relacionados con ellas (derechos naturales, individualismo, estado de naturaleza y pacto social, etc.) se generalizaron y sistematizaron entre la escolástica española[66]. También fue un lugar común en la neo-escolástica (heredado, como veremos, por el primer racionalismo) la fidelidad al dogma aristotélico de que el hombre es político por naturaleza. Por lo tanto, pese a la presencia de los conceptos contractualistas

[66] Cfr. Gough, J.W., *The Social Contract*, cit., p. 68.

fundamentales, la explicación del paso del estado natural a la sociedad civil continúa siendo un tanto providencialista.

Se produce así, en el pensamiento moderno español, una rara tensión entre la fidelidad formal a los dogmas religiosos y filosóficos de la escolástica y la forja y asimilación de unos conceptos mucho más apropiados para describir la voraz realidad moderna. Así, Juan de Mariana, uno de los más audaces y libres de los jesuitas españoles[67], inicia el *De Rege et Regis Institutione*[68] reconociendo que el hombre es por naturaleza un animal sociable, para escribir, poco después,

> "Pero ¿qué habría más inhumano y feroz que el hombre si no le detuvieran las normas del derecho y el temor a los tribunales? ¿habría acaso fieras que causasen más estragos?"[69]

Más que un hombre sociable por naturaleza, Mariana parece estar describiendo, medio siglo antes, al individuo hobbesiano. En cualquier caso, y a pesar de las contradicciones en la descripción del estado de naturaleza, los derechos individuales y la capacidad legitimadora del consentimiento unánime parecen estar fuera de toda duda, cuando Mariana continúa,

> "En mi opinión, la potestad regia, en cuanto es legítima, ha sido establecida por el consentimiento de los ciudadanos."[70]

Y, haciéndose eco de la teoría del límite del poder real establecida por Vázquez de Menchaca, concluye que

> "...por ello estimo que debió ser limitado por leyes o normas que

---

[67] Elegimos a Juan de Mariana como representante del pensamiento español del siglo XVI a sabiendas de que no puede ser catalogado estrictamente como un neo-escolástico. Más bien fue un pensador "díscolo". Sin embargo, su estilo directo y sus avanzadas ideas transmiten mejor que los representantes más "ortodoxos" algunas de las que consideramos ideas-fuerza de la Escuela de Salamanca. Por otro lado, en Mariana la teoría del contrato aparece con la mayor rotundidad, y rodeada de unos interesantísimos tintes liberales *avant la lettre*.

[68] Todas las referencias y citas son a la versión castellana de la edición bilingüe de Luis Sánchez Agesta, *La dignidad real y la educación del rey*, Madrid, Centro de Estudios Constitucionales, 1981.

[69] *Op. cit*, p. 25.

[70] *Op. cit.*, cap. VIII, p. 93.

> se estimaron necesarias, para que el poder no se salga de sus límites en perjuicio de los que están sometidos, y degenere en tiranía."[71]

Es muy dudoso que la doctrina que estos fragmentos representan pueda considerarse una teoría política contractualista en sentido moderno. Lo que es seguro es que añaden una conceptualización y sistematicidad modernas al pactismo medieval, e inician el camino hacia el contractualismo racional, representado eminentemente por Hobbes. También es de destacar que la teoría del límite del poder real conecta más directamente con proto-liberales como Locke que con el propio Hobbes.

Lo que se puede entender como un fracaso de la neo-escolástica española es que no logró construir un argumento puramente racional que conectase las ideas de estado natural, individualismo y derechos individuales, sociedad y poder. Ahora bien, se trata de un fracaso sólo aparente si se tiene en cuenta que tal construcción no fue ni siquiera intentada. La conexión entre esos conceptos nucleares del pensamiento político moderno se realizó, con toda intención, en el marco de un sistema *cuasi*-teológico, que remitía, en última instancia, a la Razón Divina como ordenadora del mundo natural y social. Si esto se considera un demérito de la escuela, recuérdese que el ponderado Locke avanza muy poco respecto a ella[72].

Hugo Grocio intentó esta conexión, apelando a un derecho natural racional. Las reglas del derecho natural, aunque reflejo del orden divino de la creación, están en una conexión necesaria con la razón humana, de modo que la mera razón, a través de las reglas naturales que infaliblemente descubre, es fundamento suficiente de la unidad política[73]. En lo que se refiere al principio de asociación, Grocio trató de evitar la apelación directa a la providencia o la "naturaleza" y recuperó el concepto estoico de *oikeiosis* (sentimiento de

---

[71] *Ibidem.*

[72] Sobre el "teocentrismo" de Locke puede verse Gauthier, D., "Why Ought One Obey God? Reflexions on Hobbes and Locke", *Canadian Journal of Philosophy*, 7, 1977, p. 425-446 (reimpreso en *Moral Dealing*, cit, p. 24-44); Dunn, J., *Rethinking Modern Political Theory*, Cambridge, Cambridge University Press, 1986, cap. 3., y la "Introducción" de W. von Leiden a su edición de los *Essays on the Law of Nature* de Locke (Oxford, Clarendon, 1988).

[73] Cfr. Grocio, H., *De Iure Belli ac Pacis*, Madrid, Centro de Estudios Constitucionales, 1987 (ed. de P. Merino Gómez), Libro I, cap. I, IX.

pertenencia a una misma especie), expresión de la sociabilidad humana[74]. Grocio puede considerarse, así, como primer representante del contractualismo político moderno. Sin embargo, la idea de sociabilidad no casa exactamente con los caracteres del contractualismo estricto que enumerábamos en el punto anterior. Es un apoyo del que muchos contractualistas se valen, sin percibir que la sociabilidad disminuye el contenido del contractualismo, pues, si los hombres son sociables por naturaleza, el papel del contrato es simplemente seleccionar el marco institucional adecuado para hacer efectiva esa disposición natural, quedando la sociedad como un "bien en sí", independiente de las reglas que la gobiernen. Sin embargo, supuestos individuos a-sociales, el contrato que les permite alcanzar mejor sus objetivos, incluye a la sociedad como instrumento. Y esa concepción inicial de la sociedad expresa el radical individualismo moderno en que se apoya el contractualismo.

c) La revolución hobbesiana y sus consecuencias.-

En el ambiente predominantemente escolástico y teológico que dominaba —y habría de seguir dominando— el pensamiento político del siglo XVI y comienzos del XVII, las obras de Hobbes fueron completamente revolucionarias. La dimensión del racionalismo que adoptaron (apoyado en las ciencias naturales, en vez de en los conceptos de "recta razón" o "razón natural"), su compromiso con un individualismo político y ontológico radical y la terrible coherencia con sus postulados iniciales, lo convierten en el filósofo político más relevante de la época, y en punto de referencia de la tradición contractualista desde entonces en adelante. Dedicaremos a la obra de Hobbes el epígrafe siguiente,por lo que aquí sólo queremos dejar constancia de que representó un punto de inflexión en la línea evolutiva de las ideas pactistas y contractualistas políticas diseminadas durante la Edad Media y maduradas al comienzo de la modernidad. Parecería que a la altura de los tiempos de Hobbes, la mentalidad moderna estaba preparada para escuchar lo que desde hacía decenios estaba ya implícito en los trabajos de juristas y filósofos. Sin embargo, las reacciones ante la obra del inglés demuestran que esto no era así. A pesar de los antecedentes, Hobbes se adelantó a su tiempo y sólo algunos "herejes"[75]

[74] Cfr. Grocio, H., *De Iure Belli ac Pacis*, cit., Prolegomenos, 6.

[75] Por ejemplo, Spinoza, a quien nos referiremos inmediatamente.

supieron o pudieron percibir la potencia y radicalidad de su pensamiento.

Pues, en efecto, Hobbes desnuda a los conceptos desarrollados por la escolástica española y la escuela del derecho natural de sus restos aristotélico-tomistas o teológicos y, tomándolos bajo una perspectiva carente de prejuicios, construye el paradigma del contractualismo político: una teoría que parte de supuestos anti-sociales (individualismo, egoísmo, guerra de todos contra todos) para construir un argumento racional que justifica la sociedad y la sumisión al poder político. Tanto la unión de los hombre en una sociedad como la autorización del soberano dependen de un pacto, de un verdadero contrato entre agentes independientes a los que no une ninguna otra relación mutua. Hobbes inaugura, así, la era de las teorías políticas del contrato hipotético[76].

Nuestra opinión es que el contractualismo hipotético estaba ya implícito en las teorías del consentimiento político tácito, como la de Juan de Mariana, o en el contractualismo "ius-naturalista" de Grocio. El mérito de Hobbes consistió en desembarazar la teoría política propiamente dicha de los fragmentos de teoría moral y teología que se adherían a ella. Su diáfano discurso fue considerado impío por sus contemporáneos, pero, a cambio, alcanza con plena vigencia nuestro siglo. Aquél discurso suponía —pese que hoy lo juzguemos repleto de innecesarias referencias religiosas— el primer argumento normativo enteramente independiente del concepto de Dios desde el convencionalismo griego.

El contractualismo hobbesiano generó más rechazo que admiración. Entre quienes lo rechazaron, ha pasado a la historia Locke, como prototipo de un contractualismo político anti-hobbesiano. Entre quienes lo admiraron, destaca Baruch Spinoza. Es admitido que Spinoza conoció a Hobbes. Si no hubiera sido así, estaríamos ante una coincidencia sin precedentes, porque los capítulos XVI y XVII del *Tratado teológico-político* reproducen casi literalmente algunas de las ideas básicas de Hobbes. Esta coincidencia nos exime de la tarea de exponer el contractualismo de Spinoza, pues basta decir que su concepción de los derechos naturales y el estado de naturaleza, así como su

---

[76] Tales teorías descansan en el siguiente esquema argumental (seguimos a G. Kavka en *Hobbesian Moral and Political Theory*, cit, p. 398):

(1) Si las personas fuesen racionales y estuviesen en tales y tales circunstancias, elegirían o acordarían un ordenamiento social de cierto tipo.

(2) Luego, las personas que realmente viven bajo un ordenamiento social de ese tipo deben obedecer las reglas de ese ordenamiento y a los oficiales designados para imponerlas coactivamente.

idea sobre el fundamento y cometido del pacto, coinciden con la del filósofo de Malmesbury[77].

Spinoza es original, sin embargo, al tratar el contenido del contrato. Mucho más preocupado por la tolerancia y la libertad que Hobbes, Spinoza cree que nunca la razón puede recomendar ceder completamente todo el derecho o poder natural del individuo, sino sólo una parte —la que convenga según la utilidad para cada uno— y, en todo caso, nunca "hasta el punto de dejar de ser hombre"[78]. Spinoza deja un lugar para lo que hoy denominaríamos "derechos humanos" (especialmente las libertades de credo y expresión, como es sabido) dentro de la sociedad civil. Tales derechos, que proceden del poder absoluto que los hombres gozan en el estado natural, no serían objeto de renuncia en el pacto sencillamente porque sería innecesario para alcanzar el objetivo del acuerdo inicial: la paz, la seguridad y el establecimiento de ayuda mutua que previniese la miseria.

Spinoza supone, por tanto, un desarrollo del contractualismo político hobbesiano hacia una mayor tolerancia y libertad. Bajo nuestro concepto —y a la luz de los más recientes análisis de la obra de Hobbes, que encuentran que sus postulados no son suficientes para explicar el absolutismo— el contractualismo político de Spinoza representa, de modo más ajustado que el hobbesiano, una reconstrucción racional del pacto fundante tal como se deriva de los supuestos (comunes) sobre racionalidad y derechos individuales[79].

Frente a Spinoza, Locke encarna un modo opuesto de solucionar el problema de la intolerancia y el absolutismo asociados al Leviathan. Si Spinoza desarrolla un pensamiento político liberal y tolerante dentro del marco contractualista hobbesiano, Locke renuncia por completo a ese marco —e incluso al contractualismo mismo— y retorna a una visión teocéntrica de la

---

[77] Cfr. Spinoza, *Tratado teológico-político*, Madrid, Alianza, 1986 (ed. de Atilano Dominguez), cap. XVI, pp. 332-333.

[78] Spinoza, *Op. cit.*, cap. XVII, p. 350.

[79] Spinoza, además, al eliminar explícitamente la razón en la configuración del derecho natural de cada hombre (que deriva directamente de sus deseos y capacidades) tenía un camino filosófico expedito para afrontar una reconstrucción contractual de la moralidad, tal como hizo con la comunidad política. Si hubiera afrontado ese reto, se habría convertido en el primer contractualista moral. No obstante, esto es sólo una hipótesis, porque el proyecto ético de Spinoza nada tenía que ver, como sabemos, con el contractualismo. Una referencia al origen de la comunidad moral se puede ver en el escolio de la proposición XVIII del libro IV de la *Ética* (p. 271-72 de la ed. de Vidal Peña, Madrid, Alianza, 1987).

normatividad social y política[80], basada en una moral individual teísta. Como ha señalado Dunn, "hay un resto de primitivismo en la política de Locke"[81].

Ciertamente, Locke retoma el lenguaje del "consentimiento" para explicar el origen de la sociedad civil y la legitimidad del gobierno. Entiende este consentimiento más como una confianza constante (y una aceptación de la institución básica de la propiedad) que como un contrato originario[82]. Locke es contractualista sólo en un sentido figurado: cree que la confianza en el respeto a los contratos es la base de la sociedad. Pero ese respeto no se debe a un pacto originario, sino que viene impuesto por la ley natural, que cada hombre tiene capacidad de reconocer directamente, mediante el uso de su razón. La importancia que Locke concede a la ley natural en el mantenimiento de la sociedad se aprecia en el siguiente pasaje del *Segundo ensayo sobre la ley natural*:

> "Sin esta ley, los legisladores podrían quizá por la fuerza y las armas, compeler a la multitud, pero no obligarlos realmente. Sin la ley natural, desaparece también la otra base de la sociedad, esto es, la confianza en el cumplimiento de los contratos, porque no se puede esperar que un hombre se sienta obligado por una promesa cuando le surjan mejores oportunidades, a menos que la obligación de mantener promesas se derive de una ley natural, y no de la voluntad humana."[83]

Es evidente que Locke no concibe el contractualismo sino como decisionismo (es decir, como convencionalismo) y, al tomar a Hobbes como su punto de referencia polémico, todo su afán consiste en demostrar que ningún

---

[80] Con la claridad que le caracteriza, John Dunn, lo explica así en *Rethinking Modern Political Theory*, cit, p. 55: "El deber de la humanidad, como criaturas de Dios, de obedecer al divino creador, fue el axioma central del pensamiento de Locke. Toda la estructura de su pensamiento fue 'teocéntrica'."

[81] Dunn, J., *The Political Thought of John Locke*, Cambridge, Cambridge University Press, 1986, p. 119.

[82] En este sentido, se pueden ver los comentarios de F. Pollow, "Locke's Theory of the State" (en Ashcraft, R. (ed.), *John Locke. Critical Assessments*, Londres, Routledge, 1991, vol. III, pp. 1-13), p. 3; Dunn, J., "Consent in the Political Theory of John Locke" (en Ashcraft, R. ed., *op. cit.*, pp. 524-556), p. 537.

[83] John Locke, *Essays on the Law of Nature*, Oxford, Clarendon, 1988 (ed. de W. von Leiden), p. 118.

orden político o ético estable puede construirse desde la concepción hobbesiana de la ley natural como auto-interés. Así lo escribe en el sexto ensayo:

> "Si la fuente y origen de toda ley fuera el cuidado y preservación de uno mismo, la virtud no sería un deber, sino la conveniencia de los hombres, ni habría bien alguno salvo lo útil."[84]

Pero al radicar el origen del deber en la relación del hombre con Dios, a través del canal "razonable"[85] de la fe religiosa, Locke no necesita apelar necesariamente a una "comunidad cristiana". Las normas de la sociedad son irrelevantes mientras permitan la libertad de cultos (o lo que es lo mismo, la búsqueda individual de la salvación). En palabras de John Dunn, "la completa individualización del deber religioso vacía a la organización social y su jerarquía de todo valor excepto su contingente conveniencia"[86]. En conclusión, la vuelta atrás lockeana, su tesis de que sólo la ley moral natural (entendida como obediencia al Creador, y no como simple libertad) puede ser base de la obligación política, da lugar, finalmente, a una teoría política convencionalista. Es una "inclinación natural", derivada de sus defectos e imperfecciones, lo que lleva al hombre a buscar la compañía de otros, prestando su consentimiento para hacerse miembros de una sociedad política. No hay necesidad racional en este trayecto. Más bien, como ya señalábamos al hablar de Mariana, parece ser la providencia la que dirige los pasos hacia la sociedad. Y esta sociedad, en tanto en cuanto para el individuo sólo representa un mecanismo útil y conveniente para superar las incomodidades (y la eventual, que no necesaria, guerra[87]) del estado natural, pero no fuente de normatividad absoluta ella

---

[84] John Locke, *Op. cit.*, p. 180. La versión "positiva" de esta tesis puede ser el siguiente fragmento del cap. 2 del *Segundo tratado sobre el gobierno civil* (p. 38 de la ed. de Carlos Mellizo, Madrid, Alianza, 1990): "El estado de naturaleza tiene una ley de naturaleza que lo gobierna y que obliga a todos; y la razón, que es esa ley, enseña a toda la humanidad que quiera consultarla que, siendo todos los hombres iguales e independientes, ninguno debe dañar a otro en lo que atañe a su vida, salud, libertad o posesiones. Pues como los hombres son todos obra de un omnipotente e infinitamente sabio Hacedor, y todos siervos de un señor soberano enviado a esta mundo por orden suya para cumplir su encargo, todos son propiedad de quien los ha hecho, y han sido destinados a durar mientras e Él le plazca, y no a otro."

[85] Cfr., sobre esto, Dunn, J., *Political Obligation in its Historical Context*, Cambridge, Cambridge University Press, 1990, p. 249.

[86] Dunn, J, *Op. cit.*, p. 250.

[87] Cfr. el cap. 3 del *Segundo tratado sobre el gobierno civil*, cit.

misma, es concebida como fruto de una sabia convención humana:

> "...un acuerdo con otros hombres, según el cual todos se unen formando una comunidad, a fin de convivir los unos con los otros de una manera confortable y pacífica, disfrutando sin riesgo de sus propiedades respectivas y mejor protegidos frente a quienes no forman parte de esa comunidad."[88]

El concepto clave en la teoría lockeana del origen y mantenimiento de la sociedad política es el "consentimiento". Pero con ello no profundiza en el contractualismo, pues no se refiere a un pacto racional necesario, sino a acuerdos contingentes, ocasionados por la "debilidad humana", no requeridos por la ley natural:

> "...por virtud de esa ley [natural], él y el resto de la humanidad son una comunidad, constituyen una sociedad separada de las demás criaturas. Y si no fuera por la corrupción y maldad de hombres degenerados, *no habría necesidad de ninguna otra sociedad*, y no habría necesidad de que los hombres se separasen de esta grande y natural comunidad para reunirse, mediante acuerdos declarados, en asociaciones pequeñas y apartadas unas de otras."[89]

La ley y la razón natural son, para Locke, fundamento suficiente de toda normatividad (incluso política, pues establecen una "comunidad natural"). Esa ley asegura, antes de cualquier pacto o contrato, la existencia y legitimidad de derechos individuales de origen natural (y divino), que nadie puede ser obligado a ceder contra su voluntad. Así, los individuos "consienten", por su conveniencia, en pertenecer a una sociedad y en obedecer los mandatos del soberano en cuanto no violen sus derechos. Pero la teoría de Locke supone una sociabilidad natural y mantiene siempre el fundamento ius-naturalista clásico de las normas justas.

Locke es considerado un contractualista en gran medida por el lenguaje que emplea en los tratados —tomado de Hobbes precisamente a fin de contradecirlo— pero la reciente recuperación de sus *Ensayos sobre el derecho*

---

[88] J.Locke, *Segundo tratado sobre el gobierno civil*, cit., cap. 8, p. 111.

[89] J. Locke, *Op. cit.*, cap. 9, p. 135, subrayado mío.

natural, que iluminan enormemente los tratados, proporciona una nueva perspectiva sobre el alcance de su teoría. Se podría decir que Locke desarrolla un contractualismo "débil": caracteriza el estado de naturaleza como un estado moral (las partes del contrato son ya agentes morales)[90]; la sociedad civil aparece como un paso "conveniente", en vez de "necesario", para asegurar la paz y, sobre todo, la propiedad y los demás derechos individuales anteriores al pacto; el contenido del acuerdo parece ser únicamente la aceptación de la regla de mayoría, u otra regla legislativa democrática, y la erección de un "juez común" (pero tanto la actividad del legislador, como la de los jueces, queda sometida a leyes naturales, no contractuales), etc.

Dados estos caracteres de su doctrina, nosotros preferimos afirmar que, pese a las apariencias, Locke no pertenece propiamente a la tradición contractualista, sino que, más bien, abre un período en que la idea y el lenguaje del contrato social se dan por supuestos, y a su sombra crecen teorías políticas de difícil clasificación[91] y variable influencia en la tradición del contrato. En el concreto caso de Locke, concluimos que su teoría es, en lo político, convencionalista y, en lo moral, naturalista y teísta.

d) La ilustración: Hume, Rousseau y Kant.-

La ilustración no es un período especialmente proclive al contractualismo político. El interés de la filosofía práctica parece centrarse en otras cuestiones, y la filosofía política, jurídica y social vivirán de las rentas racionalistas hasta la llegada del historicismo y nacionalismo románticos. Se trata, así, de un momento en que el que el pacto social es una constante del pensamiento político, pero no precisamente porque sea desarrollado o debatido en profundidad, sino, todo lo contrario, porque se convierte en un supuesto subterráneo e incuestionable del mismo, una especie de dogma. Como tal, la idea del pacto es empleada sin grandes discusiones, criticada en ocasiones y, otras veces, simplemente pasada por alto. En este ambiente resalta la figura de Rousseau,

[90] Más adelante veremos cómo la idea de establecer una restricción racional en el estado de naturaleza, deudora de los límites morales establecidos por Locke, es uno de los puntos claves del argumento de *MA*. Esta notable influencia de Locke en el neo-contractualismo (más fuerte aún en Nozick) no impide que lo consideremos más bien convencionalista. Su descripción del estado de naturaleza puede ser sugerente, pero no forma parte de un argumento contractualista como tal.

[91] Aunque netamente herederas del ius-naturalismo racionalista en su mayor parte.

impulsor de una visión revolucionaria del contractualismo que habría de cuajar en Kant y proyectarse hasta nuestro siglo. Nos centraremos, por tanto, en estos autores, añadiendo una referencia a Hume, como curioso ejemplo de la generalizada influencia del contractualismo, incluso entre quienes lo denostaron.

David Hume representa canónicamente el espíritu ilustrado. Conocidos son su método empírico y su ética "sentimental". Sus teorías de la justicia, el gobierno y la propiedad son, sin embargo, convencionalistas[92]. Hume critica explícitamente el contractualismo[93], pero su convencionalismo admite, según Gauthier, una lectura contractualista[94]. Esta lectura estaría basada en el hecho de que, aunque Hume reconoce que la utilidad es el único fundamento de las normas sociales y la justicia, sin embargo no entiende esta utilidad como utilidad social total o media. La función de la utilidad en su argumento consiste, por un lado, en servir de fundamento y explicación de las instituciones sociales y, por otro, en motivar el cumplimiento individual de ciertas reglas convencionalmente aceptadas[95]. De modo que el convencionalismo de Hume es ciertamente curioso: las reglas e instituciones que ordenan nuestra sociedad son fruto de una convención, pero su normatividad no se deriva tanto del pacto que las crea, cuanto del beneficio o utilidad que su cumplimiento comporta para el individuo, una vez situado en sociedad con sus semejantes. En última instancia, pues, el fundamento de la obligatoriedad de las normas reside en la racionalidad individual, aunque ésta se expresa a través de las convenciones sociales.

---

[92] Como lo son también las virtudes "sociales" o "artificiales" que no puedan reducirse al impulso natural del egoísmo. Sobre esto cfr. *Treatise*, libro III, pt. II, sec. II, p. 488, así como el comentario de Gauthier "Artificial Virtues and the Sensible Knave", en Tweyman, S. (ed.), *David Hume. Critical Assessments*, Londres, Routledge, 1990, vol. VI, pp. 129-154.

[93] Cfr. *A Treatise of Human Nature* , libro III pt. II sec. VII, p. 535-538 (las páginas corresponden a la 2ª ed. de L.A. Selby-Bigge, revisada por P.H. Nidditch, Oxford, Oxford University Press, 1978).

[94] Cfr. Gauthier, D., "David Hume, Contractarian", en *Moral Dealing*, cit., pp. 45-76.

[95] Pues el deber de cumplir las normas convencionales no deriva de un contrato o promesa, sino "de un sentido general de interés común, que todos los miembros de la sociedad se expresan unos a otros, y que les induce a regular sus conductas mediante ciertas reglas. Observo que será en mi interés dejar a otro en posesión de sus bienes, *dado* que él actuará igual respecto a mí..." (Hume, *Treatise*, libro III, pt. II, sec. II, p. 490).

Según esta lectura, Hume podría representar, al menos en cierta medida, un contractualismo hipotético[96], y su rechazo a la teoría del contrato social debe entenderse referido al contractualismo originario (explícito y tácito)[97]. Estas afirmaciones necesitarían matizaciones, dada la interrelación entre las teorías humeanas de la justicia, el gobierno y la propiedad, y su teoría de los sentimientos morales[98]. No obstante, muestran en qué medida el propio Hume cae dentro de la tradición del contrato, por el simple hecho de conjugar justicia e interés en el marco de una filosofía empirista en lo moral y en lo político[99]. Su expreso convencionalismo esconde, como Gauthier ha creído ver, una conexión necesaria entre la esencial igualdad y debilidad de los hombres, y su interés en respetar las normas que les aseguren una vida mutuamente beneficiosa. Esta conexión necesaria permite una lectura de Hume que lo acerca al razonamiento que conduce al contractualismo moral. Tal camino no acaba de ser recorrido, sin embargo —sostiene Gauthier— debido a la moralidad "afectiva" que abraza Hume.

Rousseau, quien no se ocupó de la moralidad de modo tan profundo como Hume, ni tenía tras de sí una tradición que lo abocara al empirismo, sí refleja en sus obras una nueva orientación política, ilustrada y democrática, de la idea del contrato social —quizá la más rica y definitiva, de la que todo el contractualismo posterior es heredero. Rousseau esboza incluso un vacilante contractualismo moral, aunque sus escuetas referencias no nos autorizan a afirmar su paternidad respecto a ese enfoque del contrato.

Lamentablemente, la sutileza e intuición de Rousseau se esconde tras la complejidad de sus textos, que están llenos de contradicciones y perplejidades.

---

[96] Su criterio sería el siguiente: son legítimas las normas cuya observancia pueda esperarse de agentes racionales auto-interesados *siempre que* puedan confiar que los demás las cumplirán de igual manera. El fundamento de legitimidad de estas normas es que ofrecen un beneficio para todos y cada uno de los miembros de la sociedad en cuestión (un fundamento típicamente individualista y contractualista).

Por otro lado, la proximidad de Hume al contractualismo hipotético estaría confirmada por su afirmación (en el *Treatise*, libro II, pt. II, sec. II, p. 493) de que el estado de naturaleza es una "ficción filosófica que nunca tuvo ni podría tener realidad".

[97] En contra de esta peculiar interpretación de Hume defendida por Gauthier, cfr. Marrone, P., "Contrato e utilità: il caso di Hume", en *Filosofia Politica*, Bolonia, vol. VI, nº 2, 1992, pp. 243-271.

[98] Cfr., en este sentido, Gauthier, "David Hume, Contractarian", cit., p. 68.

[99] Cfr. Gauthier, D., *MA*, p. 308.

Así, el titubeante contractualismo moral de *El contrato social* es contrariado por su crítica a la moral convencional en el *Discurso sobre el origen y fundamentos de la desigualdad entre los hombres*; la libertad e independencia individuales que dan lugar al contrato y que deben ser preservadas en la sociedad civil, contrasta con la versión "antigua"[100] de libertad que debería caracterizar, según su punto de vista, a la sociedad civil una vez constituida; etc. Estas dificultades nos prohíben un análisis mínimamente profundo de su obra, pues éste debería abarcar demasiados textos y, además, introducir coherencia y ligazón donde ni el propio autor lo intentó. En lugar de ese análisis, nos limitaremos a los capítulos centrales del libro I de *El contrato social*, donde Rousseau expresa de modo más claro su concepción del contrato político.

El primer dato que nos indica hasta qué punto es profunda la comprensión rousseauniana del contractualismo es el título del capítulo V de este libro I: "Cómo hay que elevarse siempre a una primera convención" —el capítulo está dirigido contra cualquier defensa del despotismo basada en el consentimiento tácito prestado a los derechos adquiridos (por fuerza, por conquista, etc.). En este capítulo Rousseau resalta el papel del pacto unánime (y, por tanto, ideal) como *único* fundamento posible de la legitimidad política.

En cuanto al contenido del pacto, es un lugar común la tesis de que Rousseau representa, frente al liberalismo de Locke (e incluso Spinoza), una escuela de contractualismo democrático que, en cuanto no pone límites a la cesión de derechos de los individuos, retrocede a un concepto hobbesiano de pacto, con la única diferencia de que aquí el cesionario es "toda la comunidad" (pues el soberano no se entiende como una persona separada de ella, sino como la comunidad misma), y el pacto produce al instante "un cuerpo moral y colectivo"[101]. En vez de el absolutismo personal de Hobbes, se trataría, por tanto, de instaurar un totalitarismo asambleario, o una tiranía de la mayoría.

Al margen del resultado político del contrato, hay que resaltar el componente moral del mismo. Rousseau parte de un estado natural simplemente a-moral para, más explícitamente que ningún otro contractualista anterior, afirmar que la justicia y la moralidad sólo tienen lugar en la sociedad:

> "Este tránsito del estado de naturaleza al estado civil produce en el hombre un cambio muy notable, al sustituir en su conducta el

[100] En el sentido de "libertad de los antiguos" de B. Constant.

[101] J.J. Rousseau, *El contrato social*, Madrid, Tecnos, 1988 (ed. de María José Villaverde), cap. VI.

> instinto por la justicia, y al dar a sus acciones la moralidad de la que antes carecían [...].
> "De acuerdo con lo anterior, podríamos añadir a la adquisición del estado civil la libertad moral, lo único que hace al hombre auténticamente dueño de sí."[102]

Con todo, Rousseau resulta inconmensurable con la tradición contractualista liberal, por lo que es difícil valorar hasta qué punto ese capítulo VIII del libro primero de *El contrato social* puede considerarse un antecedente del contractualismo moral. Desde luego, su dirección difiere radicalmente de la orientación liberal e individualista del contractualismo moral. Porque Rousseau hace consistir la libertad moral en la obediencia a la voluntad general, mientras desde el punto de vista liberal nunca las "pasiones colectivas" podrán imponerse a las individuales[103]. Su visión redentora del contrato social[104] lo reduce a un mecanismo político, cuyo efecto moral no es "creador", sino "recuperador" de la dorada edad natural. Para Rousseau, la libertad moral —entendida en la sociedad donde únicamente es posible— sólo cabe en el marco de una comunidad en sentido aristotélico. La voluntad general sustituye a la autonomía individual (aunque en la retórica democrática de Rousseau, se dirá que la expresa), de modo que, si se puede decir que la comunidad política, como un todo, queda justificada por la idea de un contrato entre sus miembros —un contrato "creador" de la autonomía de la comunidad—, no sucede así con los individuos, cuya autonomía personal desaparece tras el pacto, en un intento de recuperar la comunidad natural perdida que está destinado al fracaso.

El aroma de contractualismo moral que detectamos en Rousseau es

---

[102] Rousseau, *El contrato social*, cit., cap. VIII.

[103] Tomamos esta terminología del artículo de D. Gauthier, "The Politics of Redemption" (en *Moral Dealing*, cit., pp. 77-109). En este estudio Gauthier sugiere que Rousseau fue consciente de que las "pasiones individuales" nunca podrían ser superadas por las colectivas (el patriotismo, la religión civil, etc.). Sin entrar a valorar esta sugerencia, que la evidencia textual desmiente, hay que constatar que el concepto rousseauniano de comunidad política y moral contradice por completo las bases de una moral contractual liberal.

[104] El contrato intentaría recuperar la felicidad natural perdida en una sociedad que no es "la sociedad de su tiempo", sino una utopía pseudo-escatológica. La sociedad de su tiempo representa más bien el estado de miseria moral y de esclavitud del que hay que salir. En este sentido redentor (el más coherente con el *Discurso* y con otras obras de Rousseau), el contrato no puede ser base de una moral por acuerdo, sino un medio de recuperar la virtud y la inocencia naturales perdidas. Pero, como hemos dicho antes, cualquier intento de introducir coherencia en el pensamiento de Rousseau está, probablemente, condenado al fracaso.

únicamente la forma filosófico-moral que adopta un intento casi contradictorio: el de emplear un método individualista y liberal, como el del contrato social, para justificar racionalmente el origen y mantenimiento, no de una sociedad, sino de una comunidad política orgánica. Gran parte del atractivo de los textos de Rousseau reside en la constelación de sugerencias nacidas de este disparejo matrimonio. Pero también se deriva de ahí la escasa relevancia de su teoría para la tradición liberal del contrato.

Tal vez la redención de Rousseau como teórico del contrato social liberal (una redención que él, con toda probabilidad, rechazaría) pueda encontrarse en su gran influencia sobre Kant. Si bien la filosofía práctica kantiana, en general, obvia cualquier teoría del contrato, ya que niega que la vertiente empírica de la racionalidad (que produce solamente imperativos hipotéticos) pueda tener relevancia moral, sus apuntes de filosofía política y jurídica hacen un uso muy interesante de la misma[105]. Este uso representa un avance definitivo en el progreso de la teoría del contrato hacia una cada vez mayor abstracción e idealización. En lo político, las ideas de Kant conectan más con las de Locke o Spinoza que con las de Hobbes o Rousseau, pero su método profundiza en el camino contractualista de estos últimos.

Por ello, Kant puede ser considerado sin violencia un representante del contractualismo político liberal[106]. Porque, en efecto, aunque los materiales provenían previsiblemente de Rousseau[107], Kant formuló con toda nitidez una acabada versión del contrato social liberal entendido como expediente lógico-hipotético. De un modo que hoy es familiar, Kant se refiere al contrato como a una idea regulativa. Concibe un contrato suscrito por sujetos políticos ideales ajenos a las condiciones temporales o sociales reales. Tal contrato no está en el origen de la sociedad, ni hay que suponerlo implícito en la misma (no es un

---

[105] Por otro lado, Jiménez Perona aclara que "hay una afán de continuidad, pese a las diferencias, entre la ética y la filosofía política de Kant, como se evidencia en [los] tres rasgos propios del contrato social [...] 1. Se presenta a los individuos como un *deber*, como una exigencia moral. 2. Es una *condición formal* de los demás deberes externos. 3. Es *limitativo*: instaura una limitación recíproca de las libertades para garantizar la vigencia de los derechos de todos." (*Entre el liberalismo y la socialdemocracia*, Barcelona, Anthropos, 1993, p. 42).

[106] Parte de las ideas de estos próximos párrafos proceden de la obra de Jiménez Perona, A., *Entre el liberalismo y la socialdemocracia*, arriba citada, en esp. cap. I. El texto básico del contractualismo kantiano es su opúsculo *En torno al tópico: "tal vez eso sea correcto en teoría pero no sirve para la práctica*, Madrid, Tecnos, 1986 (trad. de M.F. Pérez López y R. Rodríguez Aramayo); esp. cap. II, "De la relación entre teoría y práctica en el derecho político".

[107] Cfr. Jiménez Perona, *op. cit.*, p. 43.

hecho), simplemente actúa como idea regulativa que, por su origen puramente racional, es inalcanzable en la práctica, pero orienta la evolución de la sociedad, establece una condición formal de los deberes jurídicos y políticos[108], y permite juzgar sobre el progreso o retroceso de los cambios sociales.

Es cierto que Kant reduce el alcance de este contractualismo hipotético estrictamente al marco de los deberes externos, pero también es cierto que en sus escritos, la idea del contrato alcanza un grado de abstracción que ya sólo va a ser igualado en nuestro siglo, con la obra de John Rawls. Así puede deducirse del siguiente texto, que resume cuanto hemos comentado:

> "Pero respecto de ese contrato (llamado *contractus originarius* o *pactum sociale*), en tanto que coalición de cada voluntad particular y privada, dentro de un pueblo, para constituir una voluntad comunitaria y pública (con el fin de establecer una legislación, sin más, legítima), en modo alguno es preciso suponer que se trata de un *hecho* [...]. Por el contrario, se trata de una *mera idea* de la razón que tiene, sin embargo, su indudable realidad (práctica), a saber, la de obligar a todo legislador a que dicte sus leyes como si éstas *pudieran* haber emanado de la voluntad unida de todo un pueblo, y a que considere a cada súbdito, en la medida en que éste quiera ser ciudadano, como si hubiera expresado su acuerdo con una voluntad tal."[109]

Claramente confina Kant al contrato a su papel de *ratio cognoscendi* de los principios formales de la legislación y el Estado[110]. Estos principios[111] configuran un estado civil subordinado al fin moral individual, pero necesario, como medio, para alcanzar esos fines. De este modo se articula la filosofía moral con la filosofía política en Kant. Este nexo deja en un lugar poco

---

[108] Esta condición consiste en que han de poder emanar de una *voluntad pública*, concepto inspirado en el de "voluntad general" de Rousseau, pero distinguible por su uso puramente formal. La voluntad general expresa en Rousseau el interés de la comunidad por encima de los individuos, mientras la voluntad pública representa, para Kant, simplemente un límite formal al legislador, de modo que éste respete siempre los intereses y derechos de *todos* los ciudadanos (que legisle *como si* sus decretos pudieran haber sido aprobados por todos).

[109] I. Kant, *En torno al tópico...*, cit., pp. 36-37.

[110] Cfr. Jiménez Perona, *op. cit.*, p. 47.

[111] Básicamente, libertad, igualdad y autonomía; cfr. Kant, *En torno al tópico...*, cit., p. 27.

prominente a las ideas políticas, lo cual es una lástima, dado el grado de precisión y abstracción que alcanzaron. La revolución que, para la filosofía práctica, supuso el formalismo kantiano postergó, durante el siglo XIX, la tesis contractualista que, si ya fue denostada en su versión hobbesiana por sus consecuencias totalitarias e impías y en su versión rousseauniana por considerarla semilla revolucionaria, fue olvidada en la versión más prometedora, la kantiana.

Cuando la crítica a la modernidad ha socavado gran parte del subsuelo ontológico sobre el que Kant construyó su filosofía moral, la idea del contrato permite recuperar, en lo político (y, desde allí, en lo moral), el impulso moderno que aquella supuso.

e) Lectura de la tradición del contrato social.-

Hasta aquí hemos expuesto —breve e incompletamente, desde luego— el progreso del contractualismo desde sus orígenes hasta su versión moderna más abstracta. Nuestro interés se ha centrado en rastrear los posibles antecedentes del contractualismo moral y, en este sentido, reconocemos que nuestra búsqueda ha sido poco fructífera. Salvo algunas referencias contradictorias y de difícil interpretación en Rousseau y una remota posibilidad en Spinoza, el resto de las teorías del contrato son inequívocamente políticas.

Es posible hablar de convencionalismo moral entre los sofistas y en otros autores, como Hume. También el convencionalismo político tiene adeptos y defensores y, sobre todo, está presente en muchas teorías del contrato insuficientemente radicales. La radicalización y abstracción de los supuestos del contrato —así como del sentido de la teoría— se observa sobre todo a partir de Hobbes: en Spinoza, Rousseau y Kant. Esta evolución se erige sobre los conceptos desarrollados por la neo-escolástica española y el ius-naturalismo racionalista europeo, que hemos representado, respectivamente, en las figuras de Juan de Mariana y Hugo Grocio. Es posible observar, en definitiva, una línea evolutiva desde un mero pactismo político (que cifra la legitimidad del gobierno en un pacto expreso o tácito entre el soberano y los súbditos) hasta un criterio racional formal ("una idea de la razón", en términos kantianos) de legitimidad política. Sostenemos que esta evolución marca una línea progresiva de profundización en el sentido del contractualismo. En esta línea, la capacidad justificadora del argumento filosófico del contrato se va ampliando: las convenciones sólo legitiman la adecuación temporal a las normas aceptadas; el

pacto de sujeción legitima la obediencia política; el contractualismo moderno justifica las bases de un Estado absoluto o de una comunidad política; el contrato social kantiano establece un criterio formal universal para la legislación en cualquier república. Nuestra opinión es que la idea de un contractualismo moral supone un paso más la dirección expuesta.

No obstante, ello no mitiga nuestra conclusión: el contractualismo moral no posee, como tal, antecedentes inmediatos. Todas las teorías que hemos mencionado *suponen*, bien una moral convencional, bien una moral identificada con los preceptos de la ley natural, e independiente del contrato, cuyo contenido es político. Pese al nexo, inevitable, entre moral privada y justicia o legalidad, los argumentos políticos incluían entre sus premisas, en su mayoría, ideas tales como "derechos naturales", o intervenciones de la providencia. Incluso Kant confía en la providencia o "la naturaleza de las cosas", que conducirá a los hombres, en sus relaciones políticas, a donde quizá ellos no quieran ir[112]. Tan sólo dos autores (y el segundo, probablemente, bajo la influencia del primero) hicieron el experimento mental de eliminar cualquier intervención de la moral o la virtud, o los designios divinos, en sus razonamientos políticos. Ellos, que son Hobbes y Spinoza, tuvieron la oportunidad de acercarse más que ningún otro al contractualismo moral. Ya hemos visto que Spinoza no usó esa posibilidad, elaborando una teoría ética aparte de su discurso político. Nos toca ahora explorar su suerte en el caso de Hobbes, el análisis de cuya teoría, por ser prototipo y paradigma del contractualismo moderno nos pondrá, sea cual sea el resultado por lo que respecta al contractualismo moral, sobre la pista del neo-contractualismo liberal.

[112] Cfr. I. Kant, *En torno al tópico...*, cit., p. 60.

4. Interpretación del contractualismo de Hobbes

a) Posibilidad de un contractualismo moral en Hobbes.-

El contractualismo moral liberal de Gauthier se inspira directamente en la obra de Hobbes[113]. Pese al resultado políticamente totalitario del Leviathan, las premisas y el método hobbesianos incluyen componentes inequívocamente liberales —el radical individualismo que enfatizara C.B. Macpherson y el relativismo axiológico son sólo los más eminentes. Peter Danielson ha llegado incluso, en su comentario sobre *MA*, a establecer un paralelismo entre las respectivas relaciones de las obras de Rawls con Kant, Nozick con Locke, y Gauthier con Hobbes[114]. Ciertamente, los tres autores contemporáneos hacen un esfuerzo similar de abstracción y de reafirmación de algunas intuiciones y métodos básicos de los respectivos autores clásicos. Todos ellos emplean, en esta empresa, los avances en la comprensión del comportamiento humano aportados principal —aunque no únicamente— por la ciencia económica a través de la Teoría de la Decisión Racional. Por último, los tres apelan a su respectivo inspirador clásico como clave de interpretación de sus teorías. Hay algo chocante, sin embargo, en el hecho de apelar a Hobbes para interpretar una teoría moral (Rawls y Nozick desarrollan teorías más estrictamente políticas). Hobbes ha pasado a la historia como adalid del más contundente convencionalismo moral y proto-positivismo jurídico. Su remisión a la "espada" como garante última del orden político y moral ha facilitado la crítica; y el descrédito filosófico de Hobbes ha sido general hasta hace pocas décadas. En este epígrafe intentaremos explicar las razones por las que la admiración hacia su obra y el

---

[113] Cfr. *MA*, p. 10.

[114] Danielson, P., "The Visible Hand of Morality", *Canadian Journal of Philosophy*, vol. 18, nº 2, junio 1988, pp. 357-384; p. 358.

empleo de su método han sustituido en los últimos años, gracias al nuevo análisis de su teoría, a aquel descrédito. Las mismas razones explicarán que Hobbes haya pasado, de ejemplo paradigmático de justificación de la tiranía, a fuente inspiradora de una teoría liberal. Y el mismo análisis mostrará la posibilidad de considerar a Hobbes verdadero antecedente remoto del contractualismo moral.

Esta posibilidad radica en un aspecto del pensamiento hobbesiano que ha resaltado como nadie John Dunn, al escribir que "el problema de Hobbes es la construcción de una sociedad política a partir de un vacío ético (*ethical vacuum*). Locke nunca tuvo este problema en los dos tratados, porque su premisa central es precisamente la ausencia de tal vacío"[115]. Podríamos añadir que Locke representa, en este pasaje, a muchos otros contractualistas, como hemos señalado en el epígrafe anterior. Frente a ellos —la mayoría de los miembros de la tradición— Hobbes abre una puerta al contractualismo moral, porque parte de una presunción *contra* la moralidad: un vacío ético. Dado que el contractualismo es un método constructivista, sólo ese tipo de punto de partida vale como premisa suya[116]. El objetivo del contractualismo moral es

---

[115] Dunn, J., *The Political Thought of John Locke*, Cambridge, Cambridge University Press, 1986, p. 79.

[116] Se podría criticar aquí que ese punto de partida permite, tanto el contractualismo como el convencionalismo, de modo que no hay por qué establecer una relación privilegiada entre Hobbes y el contractualismo moral frente a su relación con el convencionalismo. Esta crítica es aparentemente acertada; pero sólo aparentemente. La causa es que, como dejamos claro arriba, el convencionalismo, aunque comparte con el contractualismo un escepticismo moral inicial, lo deriva directamente del hecho empírico del relativismo de los valores sociales. El convencionalista trata de explicar ese relativismo de un modo acorde con una visión individualista de la sociedad y su visión egoísta (directamente maximizadora) del comportamiento humano. El contractualista, por el contrario, no se centra en el dato empírico del relativismo. Más bien al contrario, el argumento contractualista suele estar formulado desde dentro de una sociedad, con pocas referencias a las sociedades exteriores, y pretende ser un argumento racional de validez universal. Frente a ese relativismo empírico, el contractualista deduce, a partir de su concepción del individuo (basada, por otro lado, en la experiencia) cómo sería un escenario pre-social hipotético, y qué comportamientos previsibles de los agentes en ese estado puede explicar el estado social subsiguiente. En la medida en que hobbes, sin referencia alguna al relativismo empírico de los valores, establece un escenario hipotético, no sólo pre-social, sino también pre-moral, abre, por primera vez, la posibilidad a un verdadero contractualismo moral, y no sólo a una especie de convencionalismo. Todo ello con independencia de que el resultado final de Hobbes sea o no realmente contractualista moral.

mostrar que de la interacción de agentes racionales a-morales —un "estado de naturaleza" respecto a la moralidad— pueden derivarse estructuras (pactos, relaciones, compromisos) con significado moral. Pues bien, este objetivo fue anticipado por Hobbes en su tratamiento del contrato social, tal como lo interpreta Gauthier:

> "Quiero enfatizar que al introducir la justicia, afirma haber establecido, y no meramente haber supuesto, la obligatoriedad de cumplir lo que resulte de nuestros pactos —y lo ha hecho sin apelar a premisa moral alguna. De este modo, Hobbes ha investido el resultado del contrato con un significado moral."[117]

Estas dos claves interpretativas de la obra de Hobbes: el "vacío ético" del estado de naturaleza y el "significado moral" del contrato, son el punto de arranque de nuestra interpretación. Ellos proyectarán una lectura que muestra, según creemos, la inusitada virtualidad moral —al menos en potencia— del contractualismo hobbesiano. Resumiremos, en primer lugar, las líneas básicas de esta interpretación, pasando a desglosar después sus distintos aspectos, más o menos al paso del propio argumento de Hobbes. Dejaremos para el punto quinto y último de este capítulo la conclusión sobre la articulación del contractualismo hobbesiano con el enfoque de la obra de Gauthier.

b) Resumen de nuestra interpretación.-

Nuestro interés en la obra de Hobbes se centra, desde luego, en el argumento contractualista mismo. Exploraremos, por tanto, la posibilidad que su obra nos ofrece para ser leída como una teoría moral contractual *incorporada* a su teoría política (pues el argumento del contrato tiene una función

[117] Gauthier, D., "Between Hobbes and Rawls", en Gauthier, D. y Sugden R. (eds.), *Rationality, Justice and the Social Contract*, Ann Arbor, The University of Michigan Press, 1993, pp. 24-39; p. 28.

principalmente política). Ello implica obviar la importancia moral de la teoría hobbesiana del derecho natural tomada independientemente; lo cual no deja de contradecir ciertos pasajes muy clásicos y representativos, aunque no determinantes, de las obras de Hobbes[118]. De todas formas, ante las contradicciones sobre el carácter moral de las leyes naturales, que en la mayoría de los textos de Hobbes son tratadas simplemente como un conjunto de "leyes psicológicas"[119], que sólo pueden recibir propiamente el nombre de *leyes* en la medida en que un poder soberano (divino o terrenal) las impone como obligatorias, creemos defendible el enfoque de nuestra lectura.

Es, de todos modos, evidente que la teoría moral de Hobbes no es explícitamente contractualista[120], pero existe fundamento suficiente, basado en la lógica del argumento y en el contenido de sus conceptos elementales, para

---

[118] Por ejemplo, el parágrafo 31, cap. III, del *De Cive*, donde leemos que "la ley (natural), por el hecho de prescribir los medios para la paz, prescribe *las buenas costumbres o virtudes*. En consecuencia se llama *moral*".

[119] Así leemos en el *De Cive*, cap. III, 33: "Las *leyes* que llamamos naturales, al no ser más que ciertas conclusiones obtenidas racionalmente acerca de lo que se ha de hacer u omitir, y dado que la ley,propia y estrictamente hablando, consiste en la palabra de aquel que con derecho ordena a otros hacer u omitir algo, no son en sentido estricto leyes, porque proceden de la naturaleza". Y, en el cap. XVII del *Leviathan*, se añade (lo que da idea del concepto de ley natural que maneja Hobbes) que "a pesar del derecho natural (que todos obedecen, cuando tienen voluntad de obedecerlo, y pueden hacerlo con seguridad), si no se hubiera erigido ningún poder, o éste no fuese lo bastante fuerte para proporcionarnos seguridad, todos los hombres confiarán en su propia fuerza y habilidad para precaverse contra otros hombres, pues así les estará permitido legítimamente."

[120] Así, Gregory Kavka la ha identificado con cierto "egoísmo de la regla" (sin referencia alguna al pacto), mientras concede al contrato social hobbesiano una función más modesta que la moral, lo que le separaría, según este autor, del contractualismo de nuestro siglo: "la versión hobbesiana de la teoría del contrato hipotético difiere de las versiones de escritores contemporáneos como John Rawls y David Gauthier al menos en un aspecto fundamental. Mientras estos escritores tratan de derivar principios de justicia social a partir del contrato hipotético, Hobbes (y la teoría hobbesiana) usa el mecanismo para un propósito más modesto: identificar las condiciones que un sistema político debe satisfacer para que sus ciudadanos o habitantes estén obligados a obedecer sus leyes, legisladores y policía" (*Hobbesian Moral and Political Theory*, cit., p. 182).

Por su lado, el mismo Gauthier reconoce que "desde el punto de vista de la teoría moral [de Hobbes] el paso crucial requiere la intervención de un *deus ex machina*" (*MA*, p. 10), con lo que indica su convicción de que Hobbes no construye una auténtica moral por acuerdo desde el vacío ético del estado de naturaleza, como sí construye su teoría de la legitimidad política desde aquellas bases.

intentar una lectura según la cual la ley moral sólo se justifica en la medida en que se supone un pacto previo (fundante a partir de premisas no-morales) engendrador de un ámbito en que la paz y la seguridad son posibles. Podemos reducir las razones en favor de esta lectura a las siguientes:

En primer lugar, la estructura misma del argumento, cuyo eje es la idea de un contrato social fuente de toda normatividad: un pacto por el que todos renuncian a la libertad o poder natural, a cambio de poder realizar con más seguridad los fines naturales individuales. En el argumento hobbesiano, la fuente de la que derivan todas las obligaciones (incluso aquellas que hacen efectivos los mandatos de la "ley natural") es ese pacto entre individuos egoístas y previsores.

En segundo lugar, la función y definición de los conceptos morales fundamentales apoya también una interpretación contractualista del deber moral. Conceptos como los de derecho natural, leyes naturales, obligación, justicia; todos ellos son definidos con cierta vacilación, porque la perspectiva que los ilumina es la de la libertad individual[121], y su función esencial es mostrar cómo abren el paso hacia un contrato político.

---

[121] Sobre los conceptos morales de Hobbes, el mejor análisis hasta ahora es el de Gauthier (*The Logic of Leviathan*, Oxford, Clarendon, 1969, cap. II). Gauthier resalta que el derecho natural, tal como lo define y emplea Hobbes, no puede considerarse tal "derecho" en sentido estricto (pues no conlleva una obligación correlativa). Se trata de la libertad individual de acuerdo con la razón. El concepto de "ley natural" se introduce como límite de esa libertad natural, esto es, del derecho natural. Luego la ley natural "deroga" parte del derecho natural. Sin embargo, las leyes naturales también son dictados de la recta razón (individual), lo que hace imposible definir la obligación en términos de leyes naturales, pues sería tanto como defender que un ser libre puede renunciar a parte de su libertad (no otra cosa es una "obligación") y no puede, por el contrario, recuperarla más, lo cual es inconcebible desde la antropología hobbesiana (Cfr. Gauthier, *op. cit.*, p. 44). Según la concepción hobbesiana de la obligación, ésta sólo puede derivarse de un pacto, nunca de los propios actos del agente que, en la medida en que permanece independiente de los demás, conserva intacto su derecho o libertad natural. De este modo, Gauthier muestra que Hobbes configura una estructura conceptual que aboca a un enfoque contractual de la obligación moral (aunque tal vez ni siquiera el propio Hobbes fue consciente de ello): pese a que las leyes naturales, dictadas por la razón, limitan la libertad natural, esta limitación sólo puede hacerse efectiva mediante la apelación a un pacto. Posteriormente, la justicia se define como la conformidad con los pactos realizados. Se produce, así, una especie de torrente que, desde una estructura conceptual individualista en la que prima la libertad, arrastra inevitablemente hacia una concepción contractual de la justicia y, como argumentaremos, la moral.

Se puede añadir que los conceptos morales fundamentales tienen un origen individualista (de donde se deriva su necesaria vocación contractual), y son, en parte, dependientes de la teoría normativa hipotética que Hobbes va a desarrollar[122]. Están, por tanto, subordinados a la paz y la seguridad, fines colectivos necesarios desde el punto de vista individual-interesado. Su estatuto es complejo: constituyen los elementos conceptuales del contrato y, por tanto, son hipotéticamente anteriores al mismo; pero a la vez, dependen del pacto para su eficacia y están lógicamente subordinados a él (pues, incluso en la razón individual, sólo surgen por referencia a un contrato posible). En definitiva, no hay en Hobbes una fácil deducción de obligaciones morales a partir de la ley natural racional *per se*, y ello constituye una razón en favor de nuestro enfoque.

Por tanto, según nuestra interpretación, sólo el pacto social (bien en acto, bien como posibilidad) permite que entren en juego los conceptos y sentimientos morales propiamente dichos (que obligan tanto en *foro interno* como en *foro externo*). Sólo la idea de ese pacto y la consiguiente posibilidad de la paz conducen a la razón individual al conocimiento de las leyes naturales, que así comienzan a obligar en *foro interno* antes (o independientemente) de la fundación del estado social, siendo su condición de posibilidad.

Dado que todo individuo está naturalmente determinado a perseguir su propia conservación y felicidad, y que la razón recomienda a todos el contrato como único camino posible para lograr ese objetivo, se puede decir que las leyes naturales, en cuanto leyes morales (es decir, en cuanto dependientes de la idea del pacto), representan una cierta objetividad ética[123]. Esta objetividad

---

[122] Cfr. en este sentido, Kavka, G., *op. cit.*, p. 290.

[123] La forma de este párrafo —y, en parte, el fondo— refleja la opinión de Hampton en el siguiente texto: "acepto en gran parte la interpretación tradicional de la teoría moral de Hobbes, pero de un modo [...] que le permite *cierta objetividad ética*, aunque no del tipo que Kant o Aristóteles intentaron atribuir a los deberes morales. Además, aunque yo no esta clase de teoría ética, me comprometo a defenderla, en nombre de Hobbes, de la acusación de ser sólo una teoría

no concierne a la teoría del valor, que es básicamente subjetivista, sino al establecimiento de una estructura normativa que permite distinciones morales (justo/injusto, correcto/incorrecto) y por referencia a la cual se establece el contenido de la virtud.

A nuestro juicio, la tensión que se crea por la convivencia del subjetivismo axiológico y el objetivismo normativo, es un rasgo característico del contractualismo moral.

Nuestra interpretación sigue sólo en parte a Gauthier[124], tomando algunas ideas y análisis de sus estudios. A su vez, creemos que Gauthier (y también otros comentaristas de Hobbes) se inspira en la interpretación de Hobbes ofrecida por Baier en *The Moral Point of View*[125], especialmente en lo relativo a la idea de que únicamente en el ámbito de una sociedad tienen lugar las distinciones morales[126]. También nos apoyaremos en los trabajos de Kavka y, especialmente, Hampton. No obstante, debemos asumir la entera responsabilidad por las violencias que eventualmente hagamos a los textos de Hobbes. Nuestra lectura no busca ni la literalidad ni la "ortodoxia" (si existe), sino explorar las posibilidades, para la teoría ética, de la estructura lógica del contrato hobbesiano tal como ha sido revelada por la crítica contemporánea. Esta exploración supone una reinterpretación del argumento central de Hobbes —introduciendo componentes o puntos de vista que el autor no introdujo o que resaltó insuficientemente— manteniendo, sin embargo, la fidelidad a su estructura, para mostrar que si la hobbesiana no puede considerarse una teoría moral contractual no es por su diseño, sino por ciertos errores materiales o condicionamientos históricos. Con ello pretendemos probar que el aparato conceptual y lógico necesario para dar a luz el contractualismo moral se

---

descriptiva, y no una verdadera teoría *moral*" (*Hobbes and The Social Contract Tradition*, Cambridge, Cambridge U.P., 1986, p. 33).

[124] Si bien en algunos artículos, y en *MA*, Gauthier es proclive a una interpretación contractualista de la moral hobbesiana, no lo es en su clásico estudio *The Logic of Leviathan*.

[125] Ithaca, Cornell University Press, 1958.

[126] Cfr. Baier, K., *op. cit.*, p. 237 y 239.

encontraba ya, *in nuce*, en la obra de Hobbes, aunque sólo se haya desarrollado tres siglos y medio después.

c) El punto de partida del argumento.-

Los primeros capítulos del Leviathan están dedicados a la biología, la antropología y la psicología[127]. Las tesis que allí defiende Hobbes (y las que implícitamente admite) constituyen el punto de partida de su argumento contractualista.

A la vista del carácter pretendidamente empírico del contenido de esos capítulos, Kavka habla de una "teoría descriptiva" de la naturaleza humana y del desarrollo probable de la interacción natural[128]. Esta teoría se desarrollaría en paralelo a su "teoría normativa", que comprendería la teoría moral y las implicaciones normativo-políticas del contrato hipotético. Según Kavka, el ámbito descriptivo se extendería tanto a la teoría sobre la racionalidad y la acción humanas (incluyendo la teoría empírica de los valores subjetivos), como a la teoría hipotética —pero empírica en cuanto está basada en probabilidades razonables, a partir de las suposiciones empíricas referentes a los seres humanos, sus patrones de interacción y su medio[129]— sobre el conflicto en el estado de naturaleza y su posible superación mediante el acuerdo.

Nosotros establecemos una distinción entre la descripción de las partes (y su interacción natural), y el debate sobre el contrato, mecanismo sugerido por la razón para mejorar el resultado probable de la interacción natural y, por tanto, cargado con cierto componente normativo. En este sentido, diferimos de Kavka, pues no incluiríamos la hipótesis del pacto en la "parte descriptiva" de

[127] Incluyendo teorías de la percepción, el conocimiento y el comportamiento humano.

[128] Frente a esta interpretación, Cfr. Hampton, J., *op. cit*, p. 46.

[129] Cfr. Kavka, G., *op. cit.*, p. 173.

la teoría. No obstante, sí aceptamos la distinción, dentro de esa parte, entre un componente puramente descriptivo (la antropología) y otro hipotético-deductivo (la interacción y el conflicto natural). Ambos componentes sucesivos configuran el "estado de naturaleza", y corresponden, *grosso modo*, a la "caracterización de las partes" en la teoría de Gauthier.

La antropología hobbesiana puede resumirse, según Kavka, en seis rasgos esenciales, a saber: egoísmo, aversión a la muerte y al riesgo, interés por la propia reputación, previsión (se trata de individuos previsores de sus intereses a largo plazo), conflictividad de los deseos (que implica que, frecuentemente, no puedan ser satisfechos a la vez los deseos de varias personas), y aproximada igualdad intelectual y física (al menos en cuanto que todos son igualmente vulnerables)[130].

Esta antropología supone modos de pensar muy adelantados a su época. David Gauthier ha precisado el sentido en que esto es así al determinar los "tres dogmas" que, provenientes de la economía, establecen la problemática de la teoría moral moderna. Estos tres dogmas serían —por sólo enunciarlos— que el valor es utilidad subjetiva, que la racionalidad es maximización, y que las personas que interactúan no tienen interés en los intereses ajenos (auto-interés)[131]. Aunque muchas teorías morales rechazan alguno de estos dogmas, lo más valiente sería aceptarlos e intentar fijar el lugar de la moralidad a partir de ellos. Esta audaz postura, abrazada por el neo-contractualismo, tiene su antecedente en Hobbes. La caracterización hobbesiana del estado de naturaleza, efectivamente, se basa en el subjetivismo y relativismo axiológico[132], en una

---

[130] Cfr. Kavka, G., *op. cit.*, pp. 33-34.

[131] Cfr. Gauthier, D., "Thomas Hobbes, Moral Theorist", en *Moral Dealing*, cit., pp. 11-23; p. 11.

[132] *Leviathan*, libro I, cap. VI: "Pero cualquier cosa que sea el objeto del apetito o deseo de un hombre, eso es lo que él, por su parte, llama *Bueno*; y el objeto de su odio y aversión, *Malo*; y el de su desprecio, *Vil* o *Despreciable*. Porque estas palabras, "bueno", "malo", "despreciable", se usan siempre en relación a la persona que las emplea, sin que haya nada simple y absolutamente así, ni ninguna regla común de lo bueno y lo malo que se pueda extraer de la naturaleza de los objetos mismos, sino de la persona del hombre (donde no hay ninguna sociedad)..."

Leviathan, libro I, cap. XV: "*Bueno* y *Malo* son nombres que indican nuestros apetitos

visión instrumental de la razón[133] y en un radical egoísmo[134] (que incluso va más allá de los límites del dogma económico del auto-interés).

La posibilidad de leer la antropología hobbesiana en estos términos económico-individualistas es credencial suficiente para defender una interpretación contractualista de su teoría moral. Por que, a partir de semejantes premisas ¿qué otro camino, salvo el pacto (real o hipotético) entre los individuos, cabe elegir como fundamento de un criterio normativo intersubjetivo?, ¿qué otra razón, salvo la aceptabilidad general, puede legitimar las normas, sean éstas políticas, jurídicas, o morales?

Sin embargo, con estas cuestiones estamos yendo demasiado lejos, pues lo primero es mostrar que, efectivamente, el estado de naturaleza descrito por Hobbes incorpora los dogmas citados *sin incluir premisas morales*. Tales premisas, que amenazan bajo la forma de "ley natural", echarían por tierra la interpretación contractualista de la moral hobbesiana, pues supondrían un fundamento no contractual de la obligación moral. Respecto a esta amenaza hay que decir que, incluso en el libro de Hobbes más proclive a aceptar un contenido imperativo en la ley natural, el *De Cive*, se establece con claridad que el derecho natural no puede ser otra cosa que el ejercicio prudente de la razón, y éste aconseja la auto-defensa por los medios más adecuados[135]. Por

---

y aversiones que, según las diferentes naturalezas, usos y doctrinas de los hombres, son diferentes..."

[133] *Leviathan*, libro I, cap. XV: "...y los distintos hombres, no sólo difieren en lo que sienten que es agradable o desagradable al gusto, olfato, oído, tacto o vista, sino también en lo que es conforme o contrario a la razón en las acciones de la vida..."

[134] *Leviathan*, libro I, cap. XIII: "Así que en la naturaleza de los hombre encontramos tres causas principales de disputa. Primera, la competencia; segunda, la desconfianza; tercera, la gloria.

"La primera hace que los hombres invadan por lo que esperan ganar; la segunda, por seguridad; y la tercera, por conseguir reputación..."

[135] *De Cive*, cap. I, 7 y 8, p. 18. El texto de Hobbes dice: "lo que no va contra la recta razón todos dicen que está hecho justamente y con derecho (*Iure*). Por el término derecho (*Iuris*) no se significa otra cosa que la libertad que todo el mundo tiene para usar de sus facultades naturales según la recta razón. Y de este modo, el primer fundamento del derecho natural consiste en que *el hombre proteja, en cuanto pueda, su vida y sus miembros*.

"Como si se niega el derecho a los medios necesarios, el derecho al fin resulta vano, de

tanto, el derecho natural es más bien una "licencia" o "libertad" natural, que no lleva aparejada obligación alguna. Esta libertad se ejerce según la recta razón individual, o prudencia. El contenido moral de la racionalidad avendrá únicamente cuando la razón individual sea sustituida por una razón pública, que ni existe, ni puede existir, por definición, en el estado natural. No obstante, Hobbes habla de las leyes de la naturaleza como "recomendaciones" de la razón (individual) que indican el camino para la paz y, en ese sentido, restringen la libertad natural. Las leyes naturales podrían considerarse, entonces, como límites internos (morales) a la libertad. Comentaremos esta posibilidad más abajo, a modo de excurso. Pues, en rigor, el estado de naturaleza se construye sin referencia a las leyes naturales, que sólo intervienen efectivamente en el momento del pacto.

La descripción de los individuos fuera de la sociedad aboca a Hobbes a la conclusión de que el estado natural es un estado de guerra. Sin embargo, no hay por qué suponer que el egoísmo y la racionalidad instrumental al servicio de los fines individuales conduzcan necesariamente al conflicto. Podrían conducir, simplemente, a la independencia de los agentes (en un territorio suficientemente grande y con recursos abundantes). En cualquier caso, los críticos contemporáneos, iluminados por los análisis de la cooperación debidos a la Teoría de Juegos[136], han estudiado en profundidad las posibles causas del conflicto en el estado de naturaleza hobbesiano y el posible "error" de Hobbes en su formulación.

Tal como Hobbes describe la situación en el estado de naturaleza caben dos posibles fuentes del conflicto: la racionalidad de los agentes o, precisamente todo lo contrario, sus pasiones (es decir, su irracionalidad). Adelantemos que un argumento contractualista exitoso ha de mostrar que el conflicto surge naturalmente como consecuencia de una interacción racional entre individuos en una situación pre-social, pues éste es el modo de probar que la propia

---

ahí se sigue que al tener todos derecho a conservarse, todos tengan también el derecho a *usar de todos los medios y a realizar cualquier acción sin la que no podrían conservarse.*"

[136] Cfr., básicamente, Axelrod, R., *La evolución de la cooperación*, Madrid, Alianza, 1981.

racionalidad demanda la cooperación. Esto explica que Hampton, quien dedica un exhaustivo análisis a las causas del conflicto natural[137], ofrezca primero una explicación basada en la racionalidad y, ante su posible fragilidad, una segunda explicación basada en las pasiones. La explicación "racional" del conflicto plantea el estado de naturaleza como un Dilema del Prisionero (DP) (bien único, bien iterativo). Como es sabido, la mejor respuesta racional ante un DP consiste en maximizar la propia utilidad esperada para cualquier acción posible del otro jugador (en la formulación clásica del DP, confesar), lo cual es colectivamente sub-óptimo. Lo que el DP muestra es que lo "mejor" que cada uno puede hacer, es alcanzar un resultado colectivamente "peor". Aplicado al estado de naturaleza, el significado se deduce fácilmente, lo "mejor" que cada uno puede hacer, es prevenir las agresiones de los demás, incluso mediante ataques preventivos, lo cual conduce al "peor" resultado colectivo: la guerra de todos contra todos[138].

---

[137] Cfr. Hampton, J., *op. cit.*, cap. 2.

[138] Hampton pone el ejemplo siguiente (Cfr. *op. cit.*, p. 62 y ss.): imaginemos dos agentes, *A* y *B*, que interactúan en un estado de naturaleza. Cada uno de ellos se ha apoderado de un número de bienes y territorio, pero desea más (pues más bienes y territorio representan seguridad). El modo de conseguirlo es invadir el territorio del otro agente. Así cada uno tiene dos acciones posibles: invadir o no invadir, y su deliberación tiene la forma de la siguiente matriz:

		B	
		no invadir	invadir
A	no invadir	2°, 2°	4°, 1°
	invadir	1°, 4°	3°, 3°

Los números corresponden al orden de preferencias de cada jugador (el primero para el jugador *A*, el segundo para el jugador *B*), por ejemplo, el resultado de *A* invadir y *B* no invadir es *1°, 4°*, es decir, lo más preferido para *A* y lo menos preferido para *B*. El razonamiento es el siguiente: si ninguno invade, cada uno mantiene sus bienes y se mantiene, también una cierta amenaza por la presencia del otro. Si uno invade y el otro no, el invasor gana en poder y en seguridad (suponemos, además que esclaviza al invadido), por lo que preferirá este resultado a todos los demás. Si ambos invaden, ninguno gana poder y, aunque puede que mantengan sus bienes y territorio, lo harán al coste de tener que afrontar una grave amenaza; por eso, ese resultado es menos preferido que el de la mutua no-invasión. Por último, hay algo aún peor: ser invadido por sorpresa, perderlo todo y caer cautivo. Ante esta estructura de preferencias ¿Cuál será la decisión de cada agente? Veamos el caso de *A*: *A* razona que lo mejor para ambos es la situación de mutua no-invasión, pero, imaginando que, efectivamente, *B* no le invade, el resultado de sus dos posibles

Esta explicación "racional" del conflicto tiene, sin embargo, un punto débil. Ha sido puesto de manifiesto por muchos críticos, que el estado de naturaleza establece una situación de DP iterativos, en la cual no siempre es más racional la estrategia no-cooperativa. Ya demostró Axelrod, en *La evolución de la cooperación*, que la cooperación condicional arroja los mejores resultados a largo plazo entre agentes maximizadores enfrentados a un DP iterativo[139]. De este modo, Hampton concluye que "si Hobbes aceptara el argumento del DP iterativo, debería aceptar la idea de que a razón recomienda cooperar, y ello significaría que el conflicto está en función de la irracionalidad de la gente, no de su racionalidad"[140]. Parece, entonces, que la mejor explicación del conflicto se basa en las pasiones.

De hecho, los textos de Hobbes apoyan una explicación de tipo irracionalista. Su argumento es que, aunque la "recta razón" enseña cuáles son las leyes que conducen a la paz, los hombres no siguen *in foro externo* esas leyes debido a su incapacidad para sobreponerse a las pasiones (egoísmo, orgullo, soberbia, vanidad, etc.).

El mismo desafío del Tonto (*Foole*), aunque formulado aludiendo a la razón[141], está suscitado por las pasiones pues, como explica Hobbes en su

---

acciones es: no invadir: 2°, invadir: 1°; luego, si es racional, ha de invadir. Supongamos que, por el contrario, prevé una invasión de *B*, entonces, si él mismo no invade, obtendrá el resultado menos preferido: de nuevo es más racional invadir. Esto es, haga lo que haga *B* (de hecho o en las previsiones de *A*), lo más prudente para *A* es invadir. El mismo razonamiento puede aplicarse a *B*, cuyo orden de preferencias es idéntico. Así, lo racional para ambos es invadir, pero con ello alcanzan un resultado mutuamente desventajoso: ambos han de conformarse con su tercer mejor resultado, cuando la estructura de la interacción les permitía, en principio, el segundo mejor para ambos. Sin embargo, puede comprobarse que, desde el punto de vista de la racionalidad individual, este dilema es rigurosamente irresoluble. Es más, ni siquiera es tal dilema, es decir, desde el punto de vista individual cada uno está maximizando su utilidad esperada, es decir, está haciendo aquello que su racionalidad demanda. He ahí el origen racional del conflicto en el estado de naturaleza.

[139] Cfr. también Hampton, J., *op. cit.*, p. 75.

[140] *op. cit.*, p. 76.

[141] "... dado que la conservación y felicidad de cada hombre lo compromete con su propio cuidado, no puede darse razón alguna por la que un hombre no pudiera hacer lo que piense que conduce a ese fin: y,por lo tanto, hacer o no hacer, mantener o no mantener acuerdos, no iría contra la razón si condujese al propio beneficio." (*Leviathan*, cap. XV)

respuesta a este desafío, el correcto uso de la razón mostraría lo engañoso del argumento. La razón muestra que, incluso en el estado de naturaleza, es racional mantener los compromisos si una de las partes ya ha cumplido y, por ende, es racional cumplir primero pues cabe esperar el cumplimiento de la otra parte[142]. La respuesta al Tonto conduce a Hobbes a un callejón sin salida: pareciera haber intuido el razonamiento del DP iterativo y haber comprendido que, entre agentes completamente racionales, la cooperación (cumplimiento mutuo de pactos) surgiría espontáneamente en el estado de naturaleza, pues sería una demanda de la racionalidad. Entonces, sólo las pasiones pueden explicar el conflicto de hecho[143].

En cierta medida, Hobbes yerra al conceder tanta importancia a las pasiones, porque (excepto el egoísmo, que puede aceptarse como parte de la racionalidad instrumental; como una versión obsoleta del auto-interés) se trata de sentimientos marcadamente sociales que deberían haberse eliminado en la descripción de seres fuera de la sociedad. Su apelación a las pasiones (en tanto que necesarias en el individuo) es un serio obstáculo para cualquier solución al conflicto natural que no sea coactiva.

La explicación más plausible del conflicto natural no es, según Hampton, ninguna de las ofrecidas, sino una especie de mezcla entre ellas. Tal como Hampton analiza el estado de naturaleza, el cifrar el origen del conflicto en las pasiones daría como resultado un estado natural *cuasi*-lockeano, incompatible con la teoría hobbesiana de la soberanía[144]. Por otro lado, la explicación racional no se sostiene, según el argumento del DP iterativo. La solución que Hampton propone es que el conflicto surge debido a la "cortedad de miras" de algunos agentes en el estado de naturaleza. Si bien la racionalidad prescribe cumplir los pactos, porque es individualmente beneficioso a largo plazo hacerlo, hay individuos que no son capaces de otorgar suficiente peso a la

[142] Cfr. *Leviathan*, cap. XV, y Hampton, J., *op. cit*, pp. 64-65.

[143] Cfr. *Leviathan*, cap. XVII.

[144] Cfr. Hampton, J. *op. cit.*, p. 69.

utilidad futura, y entonces se comportan como si en vez de un DP iterativo, el estado de naturaleza consistiera en un sólo DP. La presencia de estos agentes "cortos de miras" provoca la prevención de todos los demás y finalmente la posibilidad de la cooperación se aleja pese a su racionalidad[145].

La explicación del conflicto basada en la "cortedad de miras" parece plausible y acorde con el argumento y el texto hobbesiano. Sin embargo, Gauthier ha propuesto una interpretación diferente[146]. Esta interpretación se basa en la respuesta de Hobbes al desafío del Tonto. Gauthier considera que, cuando Hobbes argumenta que la racionalidad prescribe cumplir los pactos incluso en el estado de naturaleza, no puede estar refiriéndose a lo que podemos llamar "efecto-reputación". Este efecto es una de las razones "a largo plazo" para cooperar en los casos de DP iterativos. Pero en el estado de naturaleza, las probabilidades de interactuar con el mismo agente serían escasas, por lo que el efecto-reputación puede descartarse. Entonces, la razón para cumplir los contratos debe ser más profunda. Esta razón no puede basarse en el auto-interés directo, pues éste recomienda justamente lo contrario: incumplir los pactos cuando sea beneficioso hacerlo. La explicación que Gauthier ofrece consiste en suponer que Hobbes adopta aquí una "visión modificada" de la racionalidad, según la cual se torna beneficioso cumplir todos los acuerdos cuando la mayoría los cumple (es decir, cuando existe la posibilidad real de entrar a formar parte de un grupo de cooperadores), aunque el cálculo auto-interesado lo desaconseje ocasionalmente.

Esta visión modificada de la racionalidad —que, como no podía ser de otra forma, recuerda la concepción "restringida" de la racionalidad de Gauthier, a la que nos referiremos en su momento— da lugar a un curioso dilema en el estado de naturaleza. Si nadie coopera, entonces es imposible que la cooperación surja (pues ningún agente se arriesgará a "ser el primero" en cooperar,

---

[145] Cfr. Hampton, J., *op. cit.*, cap 3, pp. 80-96.

[146] Esta interpretación, desconocida en *The Logic of Leviathan*, está expuesta en "Hobbes's Social Contract", en Rogers, C.A.G. y Ryan, A. (eds.), *Perspectives on Thomas Hobbes*, Oxford, Clarendon, 1988, pp. 125-152.

ante la segura explotación que sufriría). Si un grupo bastante numeroso ya interactúa cooperativamente, entonces es racional (según la visión modificada) para todos unirse a ese grupo, dados los beneficios derivados de la pertenencia al grupo. Por tanto, habría, según Gauthier, dos resultados posibles de la interacción natural: la cooperación no-coactiva entre todos los agentes o la guerra de todos contra todos. Ambas situaciones serían estables una vez dadas. El problema es que si la situación de partida es un estado de guerra, entonces la cooperación no puede surgir espontáneamente.

La interpretación de Gauthier es interesante, porque explicaría el conflicto natural sin la intervención de las pasiones. Sin embargo, no resulta coherente con la teoría hobbesiana de la soberanía ya que no daría cuenta de la permanencia del soberano: un poder absoluto, coactivo, sería necesario para forzar a los agentes (a un número significativo, al menos) a pasar del estado de no-cooperación a un sistema cooperativo, pero, una vez establecido éste, sería estable, por lo que el soberano no tendría misión alguna[147].

Vemos, por tanto, que los intentos de interpretar racionalmente el conflicto natural chocan irremediablemente con el argumento hobbesiano, que sólo parece mantenerse a base de suponer que los seres humanos están determinados por su egoísmo, sus vicios y su cortedad de miras. El enfoque hobbesiano adolece de coherencia, porque, al suponer que el conflicto puede ser superado (y que la razón muestra cómo) está contradiciendo su propia visión antropológica. Ya en este punto empieza a ser evidente que la obra de Hobbes se encamina a probar la necesidad de la soberanía absoluta, sin reparar en que la misma lógica de su argumento le desvía de ese fin.

[147] Excepto que, volviendo a lo dicho anteriormente, se reconociera que la naturaleza humana es "esclava de las pasiones", de modo que, incluso en una situación racional estable, la tendencia a incumplir los pactos en inevitable, lo mismo, entonces, que la presencia de un soberano colectivo disuasor.

d) La solución del conflicto natural.-

La salida del estado de guerra se opera mediante un pacto. No discutiremos aquí la exacta naturaleza del pacto hobbesiano[148], que se plantea en un orden exclusivamente político. Sí es importante recordar una peculiaridad del pacto hobbesiano, que lo distingue de la mayor parte de la tradición contractual hasta entonces. Se trata de que no es un acuerdo del soberano con los súbditos, sino de los súbditos entre sí. Éstos acuerdan ceder todo su derecho natural (es decir, su libertad) a un soberano (un hombre o una asamblea) cuya voluntad, de allí en adelante, representará la razón común, a la que todos deben estar y obedecer. Pero el soberano no ha "cedido" derecho alguno (pues no se ha comprometido) y, por ello, no tiene obligación alguna respecto a los súbditos.

El contenido del pacto es una cesión de derechos y la autorización de un soberano. Al ceder los derechos, cada individuo auto-limita su libertad natural; al instituir un soberano, todos aseguran el cumplimiento del compromiso cuya obligatoriedad su razón ha admitido. El compromiso así asegurado se convierte en un verdadero contrato, cuyo incumplimiento será legal y legítimamente sancionado por el soberano.

El contrato hobbesiano es hipotético. Su papel justificador de la soberanía no exige una hipótesis histórica. Es, por tanto, una instancia normativa a-temporal. Ahora bien, se trata de una instancia que, sea en la forma de promesas privadas, sea en la forma de pacto social, es el único y verdadero origen de cualesquiera obligaciones. Siguiendo a Gauthier, se puede decir que toda obligación deriva de un acto del propio sujeto obligado (un

[148] Que, por otro lado, ha dado lugar a tesis tan curiosas como la de Hampton, quien directamente niega la existencia del mismo. Esta autora cree que el estado civil es fruto de una coordinación mutuamente ventajosa de acciones de los agentes: un convenio *cuasi*-espontáneo (Cfr. *Hobbes and the Social Contract Tradition*, cit., cap. 6, pp. 132-188). Frente a esta interpretación, puede verse la defensa de Gauthier del carácter propiamente contractual del pacto hobbesiano, en "Hobbes's Social Contract", cit., p. 135 y ss.

compromiso, esto es, la renuncia a una previa libertad natural)[149]. Ésta es la esencia de la justificación proporcionada por el contrato social: si cada obligación se deriva del abandono momentáneo (mediante un convenio) de un derecho natural, el abandono completo de todo derecho comportará una obligación absoluta hacia aquél a quien se haya cedido el derecho: el soberano. Así, el contrato social es el origen de toda obligación política, y de una obligación política irresistible.

La clave del contractualismo político hobbesiano está en demostrar que la soberanía absoluta es un medio *necesario* para la paz. Dicho de otra forma, que sólo una sociedad política bajo un gobierno totalitario tendrá capacidad suficiente para superar la fuerza centrífuga, o los impulsos a-sociales, de sus individuos. Así, en la medida en que los individuos sean racionales y previsores, preferirán someterse a un soberano absoluto antes que correr el riesgo de retornar al estado de guerra[150], pues "todos conocen la maldad de la condición humana y se sabe por una sobrada experiencia lo poco que los hombres cumplen sus obligaciones en virtud de sus promesas si se suprime el castigo"[151].

Ahora bien, la percepción de Hobbes sobre la necesidad de un soberano absoluto para asegurar la paz, no es compartida por la mayoría de los críticos. Como dice Hampton, es raro ver que alguien se transforme en un absolutista convencido por el hecho de leer el *Leviathan*[152], y eso es un signo de que la justificación hobbesiana de la soberanía no es satisfactoria. El esfuerzo de la crítica reciente tiende, por tanto, a explorar las consecuencias lógicas de la

---

[149] Cfr. Gauthier, D., *The Logic of Leviathan*, cit., p. 40 y ss.

[150] Una vez más, Hampton niega la necesidad del argumento hobbesiano. Según ella, la autorización podría entenderse, a lo sumo, como un contrato de representación (*agency*), por el que los súbditos eligen un "agente" para realizar cierto trabajo. Pero siempre quedaría en manos de los súbditos la posibilidad de "despedir" a su agente (Cfr. Hampton, J., *op. cit.*, cap. 8; esp. p. 224 y ss.

[151] Hobbes, T., *De Cive*, cap. VI, 4.

[152] *Op. cit.*, p. 189.

estructura del estado de naturaleza, sin la atadura, que Hobbes parecía tener, de tener que justificar un poder político absoluto.

La posibilidad de que el resultado del pacto no fuese (o no únicamente) la institución de un soberano absoluto, sino, antes o además, la transición hacia un estado de legalidad y justicia, fue sugerida vagamente por Kurt Baier en *The Moral Point of View*[153]. Gauthier profundizó en esta línea, a partir de un análisis mucho más literal de Hobbes, en *The Logic of Leviathan*[154]. Pero es Jean Hampton quien ha ofrecido una visión más suavizada del Leviathan, a partir de la tesis que claramente expone en el capítulo séptimo de su libro:

> "Aun suponiendo que el argumento de Hobbes no fracasara por su incapacidad para demostrar la racionalidad de crear un soberano absoluto, fracasaría, no obstante, porque no puede establecer, dada su psicología, que los hombres y mujeres sean capaces de hacer lo necesario para crear un legislador que satifaciera su definición de un soberano absoluto. Esto es, si aceptamos [...] la psicología hobbesiana, veremos que el resultado de la única clase de acto de autorización que son capaces de realizar no será la institución de un soberano absoluto."[155]

La conclusión de todos estos análisis —en especial del de Hampton— es que, o bien la psicología hobbesiana ha de ser rechazada, o bien el poder instituido no puede ser un soberano absoluto. La solución de este dilema se ha convertido en un lugar común desde el libro de Gauthier: no es posible demostrar, sobre las premisas de Hobbes, que sea racional someterse a la

---

[153] Cfr. p. 239, donde Baier aduce que lo único verdaderamente necesario para el nacimiento de un sistema legal y de justicia, y para que los conceptos morales tengan sentido, es la existencia de sociedades, esto es, de modos de vida comunes, generalmente reconocidos y seguidos.

[154] Cfr. p. 173, donde Gauthier expone sus puntos de divergencia con Hobbes, entre los que destaca su idea de que, según queda establecido el estado de naturaleza y las motivaciones humanas, es posible (y necesario) limitar el derecho del soberano.

[155] Hampton, J., *op. cit.*, pp. 197-198.

soberanía absoluta. Es racional elegir un legislador y un gobernante, pero bajo condiciones limitadoras de su autoridad, y bajo la condición general de que, si el pacto inicial es incumplido, el soberano puede ser depuesto por los ciudadanos.

Sea como fuere, la solución del conflicto natural es política: la erección de un poder común —sea absoluto, como pensó Hobbes, sea limitado, como sugieren las interpretaciones contemporáneas— cuyo fin es mantener un sistema social de reglas e instituciones que "protegen" a todos los ciudadanos de un eventual regreso al estado natural. Pero el aspecto político tiene una contrapartida del lado del ciudadano: la autoridad política implica una obligación individual de obediencia. Mas no podemos olvidar que se han definido los individuos como egoístas racionales, de manera que surge la pregunta, ¿cómo puede justificarse, ante un egoísta racional, la obligación de cumplir los mandatos del soberano? Se advertirá que esta pregunta es similar a la que plantea el Tonto: Es evidentemente racional pactar la entrada en la sociedad y la autorización de un poder político, pero ello no evita que siga siendo individualmente beneficioso, en ocasiones, incumplir lo pactado. La superación de este desafío, que corre pareja con la justificación de las instituciones políticas y de la justicia misma ante un individuo racional auto-interesado, podría descansar únicamente en el argumento de la coacción. Sin embargo, como apuntábamos arriba, Hobbes ofrece trazos de una explicación más profunda, que analizaremos a continuación.

e) Razón común y moralidad.-

No se puede negar que el propósito de Hobbes es construir una justificación del Estado y el poder político. Sin embargo, es dudoso si la fuente última de la obligación política reside en el sistema legal (penal) convencionalmente creado, o proviene de una instancia normativa anterior. Si lo primero es el caso, entonces se afirmaría correctamente que Hobbes es un antecesor del

positivismo jurídico. Por el contrario, si se muestra que lo segundo es más plausible, dicha afirmación habría de ser matizada. Y, debemos reconocer que, según la mayoría de los estudiosos, lo segundo es más plausible. Tal y como Hobbes define la *ley civil*, ésta es un mandato de quien tiene el poder estatal dirigido a los súbditos, esto es, a quienes tienen la obligación de obedecer. Y esa obligación sólo puede derivarse, según hemos visto más arriba, de un acto del mismo agente obligado. Un acto que no puede ser él mismo legalmente obligatorio, sino simplemente racional; es decir, conforme a una normatividad no legal, sino de otra índole[156].

No es extraño, por tanto, que en su respuesta al Tonto, se atisbe un argumento basado en la "recta razón" individual, antes y por encima del recurso último a la "espada". Las concepciones hobbesianas de razón natural (individual), auto-interés, pacto, obligación y justicia se articulan de modo que la misma racionalidad instrumental egoísta que conduce a la guerra de todos contra todos, proporciona las bases para justificar la racionalidad del cumplimiento de los pactos. Porque para que un pacto cumpla su fin (proporcionar seguridad a quienes lo suscriben) es imprescindible que las partes se sientan *realmente* obligadas por su respectiva renuncia pública de derechos (y seguras de la obligación de la parte contraria). Si no es así, no existe tal acuerdo, sino meras palabras que no logran fin alguno. De modo que si la racionalidad individual está realmente determinada a alcanzar el fin de la seguridad y la paz, debe estarlo asimismo a *disponerse* a aceptar obligaciones como consecuencia de su renuncia a la libertad natural. Y "aceptar realmente obligaciones" significa tanto como "estar dispuesto a cumplirlas llegado el momento, sea cual sea la consecuencia"[157]. Con razón dice Hobbes al Tonto que, si bien puede no ser racional cumplir las promesas mutuas (que podrían ser no más que palabras), siempre es racional cumplir la contraprestación de un contrato cuando ya la primera parte ha sido satisfecha, porque en este caso no hay duda de la

[156] Esta interpretación está inspirada en la visión de la ley expuesta por Gauthier en "Public Reason", en *Social Philosophy and Policy*, vol. 12, nº 1 (1995), pp. 19-42; p. 33.

[157] Cfr. Gauthier, D., *Practical Reasoning*, Oxford, Clarendon, 1963, p. 188.

sinceridad de la disposición del otro agente, con lo que no hay motivo para incumplir lo que definitivamente está recomendado por la recta razón individual.

Hobbes quiere mostrar, así, que el desafío del Tonto es, efectivamente, el de un necio; porque basta plegarse a las exigencias de la racionalidad (no de una razón común, sino de la propia racionalidad auto-interesada[158]) para encontrar inmediatamente beneficioso disponerse a cumplir los pactos, de acuerdo con la tercera ley de la naturaleza[159]. Como escribe Gauthier,

> "El compromiso de quedar obligado por el acuerdo que uno mismo suscribe no es, como piensa el Tonto, un aditamento extraño y sin motivo, sino una parte esencial e intrínseca del paso de cada persona desde la naturaleza a la sociedad, un paso que todos reconocen como racional en virtud de su objetivo formal de maximizar la satisfacción global de sus fines materiales o substantivos, cualesquiera que estos sean."[160]

Es crucial este punto: que hay un motivo racional (auto-interesado) para realizar un tránsito hacia una condición social, una situación en que ha de operarse una transformación en algunas disposiciones racionales. Esta transformación expresa la idea de que no somos "máquinas de maximizar", sino seres racionales, capaces de obligarse y de actuar conforme a normas; una suposición tan débil que no hay por qué negar que Hobbes pudiera admitirla[161].

---

[158] Cfr. Gauthier, D., *The Logic of Leviathan*, cit. p. 90.

[159] Y, a la inversa, Spinoza sugiere que en un estado de naturaleza la libertad de los individuos es tal que nadie suscribiría pactos que sabe que no va a cumplir: el engaño, cuyo uso es negociar las normas legales o morales, no tendría sentido en un estado de completa libertad natural.

[160] "Between Hobbes and Rawls", cit., p. 32.

[161] De hecho, la admite bastante explícitamente en su respuesta al Tonto, en el cap. XV del *Leviathan*. Gauthier explica esa respuesta diciendo que "Hobbes dirige al Tonto, no hacia el beneficio o coste de las acciones particulares, sino hacia el beneficio o coste de las actitudes o

Así, Gauthier cree que el sentido de la autorización de un soberano o árbitro que dirima las disputas entre los individuos, consiste en erigir una "razón común" que sustituya a las racionalidades individuales, que abocan a los hombres a un estado de guerra. En los individuos, la fundación contractual de una racionalidad común, expresada en la voluntad del soberano, opera una "transformación moral"[162]. Su criterio racional para la acción, el beneficio, es sustituido por la justicia. Tras el pacto, al ceder el derecho natural al soberano, comienza la distinción entre lo que se hace con derecho y lo que se hace sin derecho, entre lo justo y lo injusto. En ese momento la moralidad aparece en el argumento hobbesiano[163], haciendo que lo racional sea lo justo, esto es, cumplir las leyes y mandatos del soberano (así como cumplir los contratos privados ya era racional, aunque moralmente neutro, en el estado de naturaleza)[164].

La institución del soberano no sólo crea un orden político y legal que puede mantenerse y reproducirse coactivamente, sino que también opera una transformación moral en los individuos al someterse voluntariamente a la razón común. Se puede establecer el siguiente paralelismo: los individuos ceden sus derechos naturales al soberano, y consienten en disfrutar de ellos ahora sólo como derechos civiles, con los límites impuestos por el deber de obediencia al soberano; de igual modo, "ceden" su racionalidad individual[165] (y el criterio del auto-interés como guía de la deliberación) para adoptar como "recta razón"

---

disposiciones" ("Between Hobbes and Rawls", cit., p. 32.).

[162] Cfr. Gauthier, D., "Between Hobbes and Rawls", cit., pp. 33, 34.

[163] Cfr. Gauthier, "Thomas Hobbes, Moral Theorist" (cit.), p. 15.

[164] Por eso, la objeción del Tonto revela un entendimiento mediocre, superado por las pasiones, ante el cual el único argumento es el recurso a la coacción; no como necesidad racional, sino como "remedio empírico" para superar la debilidad del razonamiento del Tonto (Cfr. Gauthier, D., *The Logic of Leviathan*, cit., p. 84).

[165] Escribe Gauthier en "Thomas Hobbes, Moral Theorist" (cit.), p. 21: "...al derogar cada uno su propio derecho, se ha renunciado a la razón natural como tribunal de apelación, en favor de una razón que dicta a cada hombre lo que todos deciden que es bueno."

una razón común o pública (cuyo criterio, ahora colectivo, será la optimización)[166].

Lo más destacable es que esa razón común o pública establece un límite al auto-interés (a la prudencia) que no existía en el estado de naturaleza. El fin de la razón común, que es la paz, requiere imponer restricciones a la persecución individual del auto-interés; restricciones que puede llegar incluso a exigir que los individuos arriesguen sus vidas, lo que iría directamente en contra de la razón natural individual. Sin embargo, las consideraciones de auto-interés (como la típica del Tonto) ya no se plantean tras el contrato, porque, como escribe Hobbes, "las promesas que se hacen por un bien recibido y que son pactos, son signos de la voluntad, esto es, del *último acto de deliberación*, por el cual se elimina la libertad de no cumplir"[167]. Por el contrato, la racionalidad individual *se ata a sí misma*, de una vez por todas, porque, habiendo deducido cuál es el mejor medio para alcanzar su fin, reconoce que éste medio consiste precisamente en renunciar a la deliberación natural siempre que los demás hagan lo mismo —es decir, siempre que haya esperanza de alcanzar la paz. Con ese movimiento, la "neutra" racionalidad natural se transforma en una razón moral; pero no por ello niega su fin natural (el auto-interés), sino, todo lo contrario, demuestra que las demandas del auto-interés y la moralidad se pueden reconciliar mediante el expediente del contrato[168].

Ahora bien, al preguntarse por la naturaleza de la "razón común" y de

---

[166] Sobre esta "transformación de la racionalidad", puede verse el perspicuo análisis de Gauthier en el punto V de "Hobbes's Social Contract" (cit.), donde tematiza el problema de la obligación de asistir al soberano. Aunque Hobbes no emplea este lenguaje, se puede decir que la transformación de la racionalidad individual representa el momento en que cada individuo percibe que su finalidad e intención de vivir en paz y seguridad se maximiza mediante un *compromiso* perpetuo de dar apoyo al soberano (es decir, de estar a la razón pública cuando sea necesario), incluso si la acción concreta que le sea requerida no conduce ella misma, ocasionalmente, a su propia paz y seguridad.

[167] Hobbes, T., *De Cive*, cap. I, 10; subrayado mío.

[168] Cfr., sobre esto, Kavka, G., *op. cit.*, p. 289 y, en general, la parte segunda del libro.

la "moralidad" que comporta, Gauthier y Hampton coinciden en señalar que representan una objetividad (o, mejor, intersubjetividad) convencional. Esta tesis es plausible en la medida en que la adhesión a la razón común depende de que exista en acto, o al menos como posibilidad cercana, una sociedad civil, en la que una gran parte de los individuos se adherirían de hecho a ella. Dado que estar dispuesto a alcanzar la paz cuando los demás no lo están es irracional, la razón común sólo puede establecerse sobre la base de una convención, es decir, una regularidad justificada no tanto por su contenido como por el hecho de que la mayoría la aceptan y esperan que otros la acepten igualmente. En cuanto a la moralidad, se puede decir que ésta constituye el "único conjunto dominante de convenciones, o regularidades del comportamiento, para hombres que, por perseguir ante todo su propia conservación, deben buscar la paz"[169].

Hampton ha matizado, sin embargo, la tesis de Gauthier, en un sentido que creemos esencialmente correcto, y que puede resumirse en el siguiente párrafo:

> "Para Hobbes, la moralidad tiene una base convencional, en el sentido de que es racional que los individuos realicen las acciones cooperativas dictadas por las leyes naturales —y, por tanto, produzcan la paz— sólo si otros en la sociedad las realizan también, esto es, sólo si existe una convención sobre la realización de esas acciones. Pero no creo que las leyes de la naturaleza mismas sean verdaderas por convención. Las leyes de la naturaleza describen lo que es, de hecho, necesario para que se realice la paz: la institución de ciertas convenciones que dispongan a los hombres a actuar cooperativamente. Pero la necesidad de esas convenciones para alcanzar la paz no es convencional; para Hobbes, su institución es causalmente necesaria, dada la forma en que el mundo es, para el logro de sus fines."[170]

---

[169] Gauthier, D., "Thomas Hobbes, Moral Theorist", cit., p. 16.

[170] Hampton, J., *op. cit.*, pp. 48-49.

La última parte de esta cita introduce la distinción entre el convencionalismo "formal" de las normas morales, y su no-convencionalismo esencial. Por convencionalismo formal indicamos el hecho de que las normas morales únicamente obligan en *foro externo* en el caso de que exista una "convención sobre la realización de esas acciones". El no-convencionalismo esencial se refiere a que, tanto el contenido de las leyes naturales como la necesidad de arbitrar un mecanismo convencional para hacerlas efectivas, son una necesidad racional, evidente para cualquier ser humano. La verdad de las leyes naturales, en especial la segunda, depende de que ellas establecen una conexión causal necesaria entre los fines del individuo y los medios para alcanzarlos, no de un pacto arbitrario.

Se puede decir, entonces, que, en contra de las primera afirmaciones de Gauthier, la instancia normativa que surge del contrato —la razón común— no tiene un fundamento convencional, sino contractual. Porque responde a un argumento constructivo a partir de las racionalidades individuales, pero se trata de un argumento amarrado por la necesidad derivada de las condiciones del mundo, expresadas en sus premisas.

Solucionada la dificultad con el posible convencionalismo moral incorporado en el argumento hobbesiano, surge una duda más profunda. Dijimos arriba que la sumisión de la razón natural de cada individuo a la razón común representada por el soberano y expresada en la ley civil, operaba una transformación moral en los individuos (hacía surgir la distinción entre los justo y lo injusto). Ahora bien, ¿no hay algo que chirría en esta explicación? Desde luego que sí. En principio, el curso del argumento de Hobbes parece impecable, Gauthier lo resume muy bien diciendo que "Hobbes empieza con una concepción moralmente neutra de los agentes racionales, muestra que la interacción natural sin restricciones entre tales agentes tiene una estructura similar al DP, establece el fundamento lógico que lleva a cada agente a comprometerse con los demás a aceptar restricciones mutuas en su interacción, y la razón de cada agente para disponerse a quedar comprometido por su

consentimiento. Dispuestas finalmente sus voluntades hacia la justicia, los agentes racionales se han convertido en personas morales"[171]. Pero cabe preguntar, ¿a qué precio?, ¿no se adquiere la moralidad —como sugiere Gauthier— al precio de la autonomía? Y si esto es así, ¿se puede hablar acaso de verdadera moralidad? Si la "transformación moral" consiste en internalizar las leyes naturales únicamente como mandatos del soberano, entonces, más que en justos, los seres humanos se han convertido en cautivos.

Este es un resultado descorazonador, pero es sin duda el resultado al que conduce el argumento de Hobbes, que por eso debe confiar el mantenimiento de la justicia, en última instancia, a la fuerza de la espada, y hacer depender su teoría ética de una visión relativamente clásica de las leyes de la naturaleza.

f) Excurso: la ley natural o la ética de Hobbes.-

El enfoque que hemos intentado exponer en el epígrafe anterior desecha con todo propósito una eventual teoría moral hobbesiana derivada directamente de las leyes de la naturaleza, sin el tránsito por una "transformación moral" conectada necesariamente con la idea de contrato. Este enfoque podría basarse en una frase de Gauthier: "Las leyes de la naturaleza son el fundamento de esta moralidad. Pero no son ellas mismas principios morales"[172], o en el texto de Hobbes (del cap. XIV del *Leviathan*) con que la ilustra: "Una ley de la naturaleza [...] es un precepto, o regla general, descubierta por la razón, por la cual está prohibido para un hombre hacer aquello que es destructivo para su vida, o elimina los medios para preservarla; y omitir aquello mediante lo que cree que puede ser preservada mejor".

Si la ley natural se entiende, así, como una recomendación auto-interesada, no puede considerarse como un principio moral ella misma, sino,

[171] Gauthier, D., "Between Hobbes and Rawls", cit., pp. 34-35.

[172] "Thomas Hobbes, Moral Theorist", cit., p. 15.

en todo caso, como base de una moral fundada en el auto-interés.

No obstante, las leyes naturales se plantean con la suficiente ambigüedad en los textos de Hobbes como para dar lugar a interpretaciones divergentes. Por un lado, la ciencia sobre las leyes naturales es la "verdadera filosofía moral"; las leyes son inmutables y eternas pues, aunque los deseos y apetitos de los hombres son cambiantes y distintos, su afán de alcanzar la paz es universal, y como las leyes de la naturaleza son el medio necesario para lograrla, ellas son inmutables (objetivas)[173]. No obstante, en cuanto son recomendaciones de la razón, no se pueden denominar leyes, sino teoremas o conclusiones (universalmente válidos, como las leyes de la física). Sin embargo, si se considera que también son mandatos divinos, entonces se pueden denominar propiamente leyes bajo esa luz. Por otro lado, Hobbes dice en el cap. III, 31 del *De Cive*, que "todos los autores coinciden en afirmar que la ley natural es lo mismo que la ley moral". Ello es así —y Hobbes está de acuerdo con esa opinión— porque es una verdad racional que la paz es buena, luego son buenos (virtudes morales) los medios que a ella conducen[174].

De la ley natural se dice, además, que es una obligación o una prohibición (frente al derecho natural, que es una libertad). Pensemos en esta afirmación unida a las siguientes, todas ellas contenidas o implícitas en el *Leviathan* y/o el *De Cive*: la ley natural es la verdadera ley moral; la ley natural es ley divina[175]; la ley natural es inmutable; la razón permite conocer el bien (la paz), y el contenido inmutable de la ley natural: los medios

---

[173] Cfr. Hobbes, T., *De Cive*, cap. III, 29, p. 41.

[174] No abundaremos sobre la naturaleza, posiblemente convencional, de esta verdad. Hobbes parece contentarse con afirmar (*De Cive*, cap III, 31, pp. 41-42) que "todos reconocen fácilmente como malo" el estado de guerra, y, en consecuencia, a la paz como buena. "Y, percibiendo por la razón que la paz es buena, se concluye, por la misma razón, que son buenos todos los medios necesarios para la paz y, en consecuencia, que la *modestia*, la *equidad*, la *fe*, la *humanidad*, la *misericordia*, todo lo cual hemos demostrado ser necesario para la paz, son buenas costumbres o hábitos, esto es, virtudes. Luego la ley, por el hecho de prescribir los medios para la paz, prescribe las *buenas costumbres o virtudes*. En consecuencia, se llama *moral*."

[175] Cfr. *De Cive*, cap,. IV.

necesarios para la paz.

Estas afirmaciones conjuntas han dado lugar a que la crítica considere siempre que la ética de Hobbes no se encuentra tanto en su teoría del contrato, que es una teoría política, como en la teoría de la ley natural. Modernamente, quien más ha hecho por esa interpretación ha sido Gregory Kavka, con su tesis de que la moral hobbesiana puede reducirse a un "egoísmo de la regla".

Kavka parte de un profundo análisis de la ley natural[176], cuya conclusión principal es destacar la ambigüedad de los textos de Hobbes, que unas veces parecen apuntar a un convencionalismo moral radical, mientras otras introducen ciertas distinciones morales pre-convencionales. Kavka se decanta por este segundo grupo de textos, por creer que expresan la opinión *más meditada* de Hobbes[177]. Además, Kavka es proclive a la tesis —y así la acepta un tanto acríticamente— de que la ley natural expresa deberes morales. Su problema consiste en determinar la naturaleza de tales deberes morales. Así, se ve abocado a decidir entre dos posibilidades: bien los deberes morales naturales (leyes naturales) derivan de un mandato divino; bien derivan del propio egoísmo de los agentes. Como sabemos, Kavka defiende esta segunda tesis.

Lo original de su enfoque no es tanto, sin embargo, la interpretación de la moral hobbesiana en términos de egoísmo de la regla, como la premisa sobre la que se asienta, a saber, que el modo de obligar de la ley natural, al no estar derivado del consentimiento ni de un mandato justo (como sería un mandato divino), es propia y exclusivamente moral. Kavka insiste, frente a la opinión de que el estado de naturaleza es un reino de completa libertad, donde el poder equivale al derecho y donde los términos morales dejan de tener significado, en que "*hay* moralidad en el estado de naturaleza de Hobbes, incorporada en las leyes naturales"[178]. Defiende que las leyes naturales restringen el abanico de creencias que es racional sostener (por ejemplo, estaría prohibido por la ley

[176] Cfr. *op. cit.*, pp. 338 y ss.

[177] Cfr. Kavka, G., *op. cit.*, p. 350.

[178] *Op. cit.*, p. 357.

moral creer que la ingratitud es una virtud, o un medio apropiado para la paz); que obligan realmente a cada ser racional, y que el estado de naturaleza sea un estado de libertad (es decir, de no-coacción) no es un argumento en contra de la existencia de restricciones morales en él, más bien al contrario: la moralidad natural invita a establecer restricciones convencionales para asegurar el cumplimiento de sus reglas.

No podemos negar que la interpretación de Kavka es plausible y coherente —aunque sólo de modo parcial— con la que podemos llamar "postura definitiva" de Hobbes[179]. De hecho, el contractualismo moral *podría* entenderse, en cierto modo, como un egoísmo de la regla[180]. Ambos se basan en la idea subyacente de que es verdaderamente racional ser moral, esto es, que seguir ciertos principios cooperativos es individualmente ventajoso, desde el punto de vista del auto-interés.

Sin embargo, Kavka olvida que la efectividad de la ley natural —incluso desde la perspectiva de "teorema de la razón"— *depende* de que pueda

---

[179] Si tomamos por "postura definitiva" de Hobbes la que expone en su correspondencia con el Obispo Bramhall (cinco años después de la publicación de la versión inglesa del *Leviathan*), debemos entender que las leyes de la naturaleza, en cuando emanadas de la palabra de Dios, son verdaderas reglas morales inmutables y eternas (y en esto Kavka tendría parte de razón), pero en cuanto los hombres las conocen únicamente mediante su razón natural, no son sino teoremas que nos guían hacia la paz, pero inciertos, como conclusiones de hombres particulares que son y, por lo tanto, no propiamente leyes morales. Desde este punto de vista, la confusión de Kavka es mayúscula, pues él rechaza explícitamente la tesis de Howard Warrender de que la ley natural es esencialmente un mandato divino (Cfr. Kavka, G., *op. cit.*, p. 368), con lo que —de acuerdo a la literalidad de los textos de Hobbes— cierra la puerta a una lectura moral de las mismas. Lo incoherente (también lo ambicioso) de la interpretación de Kavka es el desgajar la obligación moral incorporada en las leyes naturales de su consideración como verdaderas leyes, es decir, como mandatos de la autoridad soberana (Dios, en el estado de naturaleza). Como teoremas de la razón, las "leyes" (ahora usando impropiamente el nombre) son el fundamento de una moral convencional o contractual, pero no son ellas mismas principios morales.

[180] El principio moral del egoísmo de la regla es el siguiente, según Kavka (*op. cit.*, pp. 358-359): "Cada agente debe intentar siempre seguir aquel conjunto de reglas generales de conducta cuya aceptación (y sincera disposición de seguirlas) por su parte en todas las ocasiones produciría el mejor resultado (esperado) para él". El contractualismo moral parte de un principio semejante, sólo que lo considera un "principio racional", empírico; no un mandato moral.

Otra diferencia es que el contractualismo no restringe el auto-interés a "lo mejor para el agente que actúa". Sin embargo, esta diferencia puede soslayarse si el egoísmo de la regla interpreta el "interés del agente" al modo económico-utilitarista, como el contractualismo.

esperarse la paz, es decir, de la idea de un contrato entre todos los agentes naturales capaz de convertir las directrices de la razón en verdaderas leyes. Sólo la ley civil (post-contractual) induce en los individuos un criterio moral. Las leyes de la naturaleza son, como dice Hampton[181], meros imperativos hipotéticos y, además, condicionales[182]. En el estado de naturaleza, las ley moral puede ser conocida por los agentes racionales gracias a su capacidad de anticipar hipotéticamente el pacto social, debido a su naturaleza racional. Pero esa ley anticipada, es constituye una moralidad efectiva; no tiene capacidad de obligar. Tal vez podría decirse, con todo, que las dos primeras leyes de la naturaleza, que ordenan perseguir la paz y ceder los derechos naturales cuando los demás hagan lo mismo, sí obligan directamente en el estado natural; pero a eso respondemos que, en ese caso, las dos primera leyes no son sino expresión de la recta razón y, en realidad, tienen el mismo carácter moral —desde el punto de vista hobbesiano— que una ley física.

Precisamente esta visión a-moral del estado de naturaleza distingue a Hobbes de la mayoría de los contractualistas políticos y posibilita nuestra interpretación de su obra como un esbozo —finalmente inconcluso— del contractualismo moral.

g) El camino hacia el contractualismo moral.-

Desechada la hipótesis de una conexión directa entre racionalidad individual y moralidad a través del concepto de ley natural, hemos de retomar el argumento expuesto arriba: racionalidad individual y moralidad se conectan a través de la idea de una "razón común" y ésta tiene su fundamento en un

[181] *op. cit.*, p. 89 y ss.

[182] Se formularían así: "Si quieres la paz, haz *x*, si los demás también lo hacen". Como quiera que el objetivo de la paz es connatural a todos los seres racionales, se podrían formular como imperativos asertóricos: "Siempre que los demás hagan *x*, hazlo tú también".

contrato hipotético entre agentes racionales independientes. La legalidad y la moralidad tendrían, pues, una justificación contractual si no fuera por la duda que planteábamos al final del epígrafe e), sobre la pérdida de autonomía asociada a la "transformación moral".

Ahora debemos preguntarnos si, aceptando la interpretación que hemos desarrollado, Hobbes recorre completamente el camino hacia un contractualismo moral o si, por el contrario, fracasa en algún momento del mismo, tal como parece sugerir aquella duda.

Según nuestra interpretación, Hobbes inicia, consciente o inconscientemente, un argumento que contiene los elementos del contractualismo moral:

En primer lugar, acepta los que Gauthier ha denominado "dogmas" procedentes de la economía, que enmarcan la problemática de la filosofía moral moderna: la relatividad del valor, la racionalidad como maximización, y el auto-interés. Al aceptar esos "dogmas" individualistas, Hobbes imagina un estado de naturaleza radicalmente a-social y a-moral, de desconocida factura hasta entonces en la literatura contractualista. Un estado de naturaleza completamente vacío de ocultos y providenciales "encantamientos" normativos. Reducido a un conjunto de rasgos empíricamente humanos y racionales. Los habitantes de ese estado de naturaleza son individuos a-morales, libres y egoístas.

Salvando la distancia temporal y de contexto filosófico, teológico y político, ésas son las premisas del contractualismo moral contemporáneo, basado en la Teoría de la Decisión Racional[183]. Y, de hecho, al igual que el neo-contractualismo, Hobbes concluye su argumento justificando obligaciones

[183] No en vano dice Gauthier que, de haber contado con los recursos de la Teoría de la Decisión Racional, Hobbes habría llevado a cabo la empresa —afrontada en este siglo por Rawls y por él mismo— de demostrar que los principios morales forman parte de los principios de la racionalidad; tesis esencial del contractualismo moral (Cfr. Gauthier, "Morality, Rational Choice and Semantic Representation", en Paul, E.F., *et al.* (eds.), *The New Social Contract*, Oxford, Blackwell, 1988, pp. 173-221; p. 173).

morales sin renunciar a sus premisas sobre la racionalidad y el individualismo[184].

Su argumento prosigue mostrando que las racionalidades individuales son capaces de conformar, mediante el contrato, una razón común necesaria —de carácter no-convencional, según Hampton. Hasta aquí, asistimos a la derivación de deberes que *coinciden*, en su contenido, con una idea pre-teórica de moralidad: deberes de justicia retributiva y distributiva, deberes de reciprocidad, de respeto mutuo, etc.[185]. De hecho, la moralidad se ha identificado en ocasiones con la internalización de esa "razón común", o con la adopción un "punto de vista" neutro, o de "tercera persona".

Sin embargo, la antropología hobbesiana no permite esa internalización. La razón común *no* se incorpora en los individuos como un desarrollo o modificación de su racionalidad auto-interesada, sino que se expresa en una teoría del poder y el derecho positivo. La razón común queda atrapada en las redes de la política; el soberano es su único garante. Los individuos alienan no sólo sus derechos naturales, sino también su capacidad natural de deliberación, de modo que entran en el reino de las distinciones morales al precio de perder su autonomía. La antropología hobbesiana es más cerrada, aún, de lo que parece: la libertad y autonomía naturales son incompatibles con toda norma no-coactiva. El sometimiento a los pactos y a las virtudes que permiten la cooperación sería posible sólo tras renunciar a la libertad natural, tras un encadenamiento (voluntario, eso sí) de la razón individual a la decisión inapelable del soberano.

La crítica ha puesto de manifiesto que esta completa subordinación de la moralidad a la política —en concreto al absolutismo político— no es en modo

---

[184] Así lo recoge Gauthier al final de "Thomas Hobbes, Moral Theorist": "La teoría moral de Hobbes es un convencionalismo dual, en el que una razón convencional que se impone a la razón natural, justifica una moralidad convencional, que restringe el comportamiento natural". (p. 22).

[185] Para una completa categorización de los deberes contenidos en las leyes naturales según Hobbes, ver Kavka, G., *op. cit.*, p. 343.

alguno una necesidad lógica del argumento hobbesiano. Mas bien al contrario, la caracterización de los individuos en el estado de naturaleza hace difícil la institución de un soberano absoluto[186]. Sorprendentemente, el Leviathan sería menos amenazador si los individuos naturales no cometieran algunos errores que Hobbes les hace cometer. Entre seres perfectamente racionales y mutuamente desinteresados, un Estado tolerante y democrático, acompañado de un verdadero sentido individual y colectivo de la justicia, es el resultado más probable del contrato social.

Analizar en profundidad las causas que hicieron que hobbes no pudiese alcanzar este tipo de conclusiones —a las que, curiosamente, se acercó más Spinoza— nos llevaría demasiado lejos. Mas podemos apuntar brevemente que estas causas se reducen a dos principales, una interna a la teoría y otra "externa".

La causa interna es que Hobbes establece unos presupuestos sobre la naturaleza humana demasiado estrictos, de los que no puede escapar. Considera que las acciones humanas están completamente determinadas —dentro y fuera de la sociedad— por el imperativo de la supervivencia y la auto-preservación. En términos económicos modernos, los seres humanos son para Hobbes "máquinas de maximizar" (lo cual es coherente con el mecanicismo que abrazaba). Aunque pueden alcanzar un compromiso de futuro mutuamente ventajoso (pues el beneficio futuro se percibe racionalmente, y partimos de la base de que son racionales), no son capaces de cumplir lo pactado, pues cuando llega el tiempo de hacerlo, el beneficio inmediato derivado del incumplimiento está mucho más presente, y arrastra inevitablemente a la voluntad. Por poner un ejemplo conocido, desde el punto de vista de Hobbes, todos somos como Ulises ante la isla de las Sirenas. Sólo atándonos a nosotros mismos —y atarse, en este contexto, significa someterse voluntariamente a un soberano que haga cumplir los decretos de la razón pública, mediante la espada si es necesario— tenemos garantizado superar una tentación que sabemos tan irresistible como

[186] David Gauthier mostró, en *The Logic of Leviathan* (Cfr. p. 115 y ss.), que intentar fundar un régimen de soberanía despótica en un contrato es lógicamente inconsistente.

perniciosa.

La causa externa es simple: la obra de Hobbes está subordinada a un fin político, cual es la justificación y defensa del absolutismo monárquico[187]. Una mera cesión de derechos por parte de los súbditos no es suficiente para tal propósito. Para Gauthier, por ejemplo, la pérdida de autonomía era inevitable si de lo que se trataba era de fundar una soberanía absoluta permanente: "Hobbes *debe* abrazar una teoría del contrato social alienador para defender la soberanía absoluta permanente. Si el súbdito meramente cede sus derechos al soberano, entonces no se asegura un poder absoluto, ni permanente"[188]. Así, mientras la crítica contemporánea ha desarrollado sin esfuerzo las posibilidades "liberales" del estado de naturaleza hobbesiano, el mismo Hobbes renunció a ellas de antemano (comprometiéndose con una antropología tan inflexible como ficticia), para dirigir su argumento hacia el resultado político que consideraba preferible.

Atrapado por sus condicionantes internos y externos, el argumento de Hobbes se muestra como una justificación —no enteramente satisfactoria, por cierto— de la soberanía absoluta. En definitiva, como una teoría política típicamente moderna, es decir, ajena al problema de la felicidad (en términos religiosos, salvación) individual.

Este aspecto de la teoría hobbesiana ha facilitado una lectura "teológica" de su filosofía moral (según la cual, su ética deriva de la obligación de obedecer a Dios), separada de su psicología y su teoría política[189].

El propio Gauthier parece abrazar esta lectura cuando reconoce, en *The Logic of Leviathan*, que "el sistema 'moral' hobbesiano no es nada más que un sistema de prudencia universal o común [...]. Lo que nos impide clasificar el

---

[187] Una intención manifiesta en muchos pasajes. Pueden verse, como ejemplos, el cap. VI, 13, p. 60 del *De Cive*, o el entero cap. XVIII del *Leviathan*.

[188] Gauthier, D., "Hobbes's Social Contract", cit., p. 151.

[189] Cfr. Hampton, J., *op. cit.*, p. 30, donde cita a Warrender, H. (*The Political Philosophy of Hobbes*, Oxford, Clarendon, 1965) y Taylor A.E. ("The Ethical doctrine of Hobbes", en Brown, K. (ed.), *Hobbes Studies*, Oxford, Blackwell, 1965, pp. 35-55) como defensores de esta lectura "tradicional" de la ética hobbesiana.

sistema hobbesiano como moral es el hecho de que los hombres estén necesariamente inclinados a su propia conservación o, más generalmente, a su propio beneficio. En este sentido, su psicología es destructiva para su ética"[190]. Mas también sugiere Gauthier la vía de escape, que no es otra que rechazar la psicología hobbesiana, al menos en cuanto supone que la motivación humana es exclusivamente egoísta (*selfish*)[191].

La enseñanza que extraemos de las interpretaciones analizadas, así como, básicamente, de los propios textos de Hobbes, es que en su obra se halla el germen de una justificación de la moralidad desde premisas radicalmente no-morales. Sólo se trata de un germen que posiblemente Hobbes no pudo reconocer. Es una posibilidad incorporada en su conceptualización del estado de naturaleza y en el radical individualismo de su teoría, pero que no eclosiona debido, probablemente, al ineludible peso teológico de la moral y a sus preocupaciones y opiniones políticas personales.

Pero la estructura del argumento conduciría, bajo unos presupuestos antropológicos ligeramente diferentes —y más realistas, como ha demostrado la Teoría de la Decisión— a la justificación de una moral por acuerdo.

Frente al resto de la tradición del contrato —cada uno de cuyos representantes añade, desde luego, elementos nuevos y vitalizadores del argumento—, Hobbes es el gran revolucionario. Su teoría es ya inequívocamente moderna y, como hemos tratado de mostrar, deja el camino franco hacia un contrato social liberal y da forma a los elementos y argumento del contractualismo moral.

---

[190] P. 98.

[191] En *The Logic of Leviathan*, cit., p. 172.

5. Sobre los antecedentes del contractualismo moral liberal

El objetivo de este capítulo era situar al contractualismo moral liberal en su tradición; explorar sus antecedentes.

Hemos encontrado que el contractualismo moral desarrollado en nuestro siglo carece de precursores directos, pero sí es legítimo verlo como la prolongación de una tradición contractualista que —especialmente en la línea que va de Hobbes y Spinoza a Kant, a través de Rousseau— tiende hacia una progresiva abstracción y universalidad. La culminación de esta línea es, por el momento, la teoría moral contractualista.

Ésta hunde, por tanto, sus raíces, en el contractualismo hobbesiano. Así es explícito, de una vez por todas, en las autorizadas palabras de Hampton:

> "Las raíces de la teoría moral de Gauthier están en el *Leviathan* de Hobbes. Afirmando que el valor es subjetivo y que la racionalidad debe definirse instrumentalmente, Hobbes concluye que los imperativos morales son hipotéticos o, con sus palabras, 'conclusiones o teoremas sobre lo que conduce a la conservación y defensa' de la humanidad. En consecuencia, la moralidad es presentada como un sistema de restricciones mutuamente beneficiosas que los individuos 'podrían acordar adoptar' a fin de promover relaciones mutuas instrumentalmente valiosas. Mas Hobbes insiste también en que sería irracional adoptar estas restricciones sin la seguridad de que un número suficiente de los demás lo hará también [...]. Sólo cuando las sanciones de un soberano puedan castigar el comportamiento contrario a la ley, encontrarán los hombres racional apoyarlas y seguirlas.
>
> [...]
>
> "Pero Gauthier se pregunta por qué habríamos de recurrir a un

> remedio político para el problema del conflicto humano cuando, tal vez, podría haber un remedio moral. Intenta defender que la naturaleza de la moralidad —incluyendo la justicia— puede definirse mediante la metodología contractualista, y que la psicología humana hace posible el comportamiento moral sin necesidad de recurrir al gobierno."[192]

Gauthier, y el contractualismo moral que él representa, al confiar en la posibilidad de establecer restricciones imparciales sin necesidad de recurrir a la estructura política de un Estado absolutista, parecen inspirarse, por cierto, en una maravillosa idea sugerida por el mismo Hobbes, en el capítulo XVII (parte II) del *Leviathan*:

> "Si pudiéramos imaginar que una gran multitud de hombres es capaz de consentir en la observancia de la justicia y las otras leyes de la naturaleza sin un poder común que mantuviera ese respeto coactivamente; podríamos también imaginar a la humanidad entera haciendo lo mismo; y, entonces, ni habría, ni sería necesario que hubiera, ningún gobierno civil, ni comunidad política alguna, porque habría paz sin sometimiento político."

Desarrollar esta idea en un marco contractualista significa tanto como defender que la disposición a la "observancia de la justicia ... sin un poder común" puede fundirse con la persecución individual del interés[193]. Y éste es precisamente el contenido básico de la teoría de David Gauthier. Un contenido

---

[192] Hampton, J., "Can We Agree on Morals", *Canadian Journal of Philosophy*, vol. 18, nº 2, junio 1988, pp. 331-356; pp. 331-332. En el mismo sentido, cfr. *Hobbes and the Social Contract Tradition*, p. 92.

[193] También se puede imaginar que esa disposición nace de un "sentimiento de simpatía" compartido por todos los humanos, pero puede probarse que, a partir de esa premisa, todo lo que cabe alcanzar es un convencionalismo moral como el de Hume —que ha influido en la orientación metodológicamente contractualista de Harman en *The Nature of Morality: An Introduction to Ethics* (Nueva York, Oxford University Press, 1977).

inspirado por la luz cercana de la obra de Baier[194], con su idea fundamental de que la moral puede entenderse como un conjunto de principios que derogan la búsqueda individual del beneficio cuando es *beneficioso* para todos que así sea. Un contenido al cual se adecua perfectamente el marco argumental contractualista, que presenta una posición original a-moral —en la que se justifican las demandas del auto-interés y una concepción instrumental de la racionalidad— para construir un sistema cooperativo enderezado al beneficio mutuo, cuyos principios pueden ser identificados con los principios de la justicia y la moralidad.

La relación entre prudencia (interés) y moralidad es iluminada, así, con la luz del contrato social. Como gráficamente dirá Gauthier, la moral aparece ante nuestro ojos a partir de una situación original especialmente concebida *contra* la moralidad. Mas no se trata de un juego de prestidigitación intelectual; se trata de mostrar la naturaleza racional de la moralidad, a través de la posibilidad de entenderla como un acuerdo hipotético entre agentes perfectamente racionales, que revela la conexión causal necesaria (como advertía Hampton refiriéndose a Hobbes) entre las demandas del interés individual y la erección de principios morales.

Por tanto, la afirmación contractualista de que las distinciones morales sólo tienen sentido en el marco de la sociedad, no quiere decir que la moral sea una simple convención social. El nexo que el contrato quiere poner de manifiesto, es una relación esencial que prueba la base racional de la moralidad. Esta base garantizará su universalidad. Pero el método (contractualista) no discute su posible origen histórico ni su ontogénesis.

Sostenemos —a partir de los análisis de los epígrafes anteriores, y teniendo en cuenta lo inmediatamente antedicho— que la tesis de que es posible entender la moralidad como un conjunto de restricciones acordadas al comportamiento natural egoísta, y explicarla mediante la hipótesis de un

[194] Cfr. los explícitos textos de Baier, K., *The Moral Point of View*, Ithaca, Cornell University Press, 1958, pp. 237, 239 y 314.

contrato ideal entre los individuos naturales, tiene su antecedente remoto en la versión del contrato social moderna que encontramos en Hobbes. Que aunque Hobbes mismo no desarrolla una teoría moral contractual, la idea de que es posible derivar la moralidad a partir del auto-interés hizo fortuna entre autores como Spinoza, Puffendorf e incluso Locke[195], y llegó a expresarse con toda su radicalidad en la obra de Baier, quien influye poderosamente en Gauthier. Que la estructura del contrato Hobbesiano (su caracterización del estado de naturaleza, la deducción de una "razón común", etc.) supone un paso revolucionario en el esfuerzo moderno de abstracción basado en la idea clásica y medieval de un pacto originario como criterio de legitimidad política. Este esfuerzo de abstracción culmina en Kant y es re-descubierto por Rawls, y establece el canon de la comprensión filosófica contemporánea del expediente justificador del contrato.

Por todo ello se puede afirmar que el contractualismo hobbesiano ha servido de modelo e inspiración para la teoría moral cuya exposición retomaremos en el capítulo siguiente.

Evidentemente hay muchas diferencias entre la posición hobbesiana y la de Gauthier[196]. Pero estas diferencias atañen precisamente a los puntos cuyo replanteamiento era necesario para poder construir una filosofía moral contractualista, apoyados, además, por los avances en la comprensión de la conducta racional.

---

[195] Cfr. Hampton, *op. cit.*, p. 57.

[196] Por citar sólo algunos ejemplos significativos: A pesar de su subjetivismo moral, Hobbes acaba por reconocer que *hay* un bien objetivo que orienta inevitable y universalmente las voluntades y los deseos humanos: la propia conservación (tanto es así que la defensa de este bien es la única causa que justifica la resistencia al soberano [cfr. *De Cive*, cap. II, 18]). En este punto, Gauthier es más radical, al unir subjetivismo y relativismo, tal como requiere su compromiso con una concepción instrumental de la racionalidad. Los fines de la razón no pueden ser determinados antes del pacto más que formalmente, como fines auto-interesados de cada individuo, pero no existe ninguna determinación substantiva, por más obvia que pudiera parecer.

Otra diferencia esencial es la capacidad de compromiso sincero que Gauthier concede a los seres humanos, y que Hobbes les niega. Frente a las "máquinas de maximizar" hobbesianas, Gauthier toma algunas características del individuo económico, pero su concepción del individuo tiene un alcance mayor.

En definitiva, el contractualismo moral entronca con el impulso moderno, máximamente representado por Kant, que convierte al hombre en legislador de la moral y la naturaleza, creador (sea en cuanto sujeto del conocimiento, sujeto moral o ciudadano) de la ciencia, la ética y la política[197]. Cuando la crítica a la modernidad ha socavado gran parte del subsuelo ontológico sobre el que Kant construyó su filosofía, la idea del contrato permite recuperar, al menos en lo político y en lo moral, el impulso moderno. La creciente comprensión de los mecanismos de la racionalidad práctica nos concede confiar en la pregnancia de las nuevas versiones de las teorías clásicas y, en concreto, en la posibilidad de encontrar una salida para el callejón del autoritarismo en el que Hobbes dejó embarrancada su idea de una "racionalidad común". Esta posibilidad está señalada en el siguiente párrafo de Gauthier, con el que queremos concluir este capítulo, para intentar analizar, en el próximo, hasta qué punto su esfuerzo de superar el hobbismo ha tenido éxito como filosofía moral:

> "Comencemos otra vez con el agente racional. Consciente del coste de una interacción libre con sus semejantes, consciente también de los beneficios potenciales de la cooperación, llega a tener conciencia de las leyes de la naturaleza. Reconoce que racionalmente debe estar dispuesto a consentir ciertas restricciones a su primitiva libertad de acción, siempre que los demás estén igualmente dispuestos, y entonces se compromete a quedar obligado por las restricciones acordadas. No renuncia a su juicio sobre el bien y el mal, pero se somete a una norma común para ese juicio, proporcionada por las leyes de la naturaleza, y esa norma le guía hacia el beneficio mutuo, renunciando a su interés individual. De este modo exhibe lo que podemos llamar, siguiendo a Rawls, un sentido de la justicia [...]. Adquiere personalidad moral mientras retiene su autoridad como agente racional, porque

[197] Cfr., sobre esto, Jiménez Perona, A., *Entre el liberalismo y la socialdemocracia*, Barcelona, Anthropos, 1993, p. 25.

internaliza las leyes de la naturaleza como exigencias racionales en su deliberación y acción. No subordina su razón a un soberano, como en Hobbes, sino que, a través de las leyes de la naturaleza, coordina su razón con la de sus semejantes."[198]

[198] Gauthier, D., "Between Hobbes and Rawls", cit., pp. 37-38.

Capítulo IV

El contrato moral

1. El contenido del contractualismo moral

Al inicio del punto cuarto del capítulo segundo, antes de comenzar nuestra discusión sobre el papel del mercado, señalamos la línea argumental de la que momentáneamente nos apartábamos. Decíamos que tras el planteamiento de las premisas teóricas y el marco axiológicamente plural que se toma como punto de partida —y que, conjuntamente, configuran los elementos de un hipotético "estado de naturaleza"—, el siguiente paso debería ser la identificación del esquema de interacción de los individuos en una "posición inicial" que incorporase las características de las premisas que habíamos definido. El argumento contractualista consiste precisamente en mostrar que la lógica de dicha interacción conduce a los agentes racionales a suscribir y cumplir un acuerdo que permita superar las imperfecciones de la interacción natural y franquear de este modo el camino hacia la maximización individual de utilidad que está paradójicamente prohibida fuera del marco de la cooperación. Con este argumento, el contractualista moral pretende fundamentar afirmaciones como la que nos servía para concluir el capítulo anterior.

El propósito del presente es, efectivamente, proseguir con el argumento contractualista propiamente dicho (en la medida en que la complejidad a que este argumento ha sido llevado por autores como Gauthier nos permite una

exposición "ordenada" del mismo). Pero antes de enlazar con el capítulo segundo, queremos plantear algunas observaciones de carácter general sobre el sentido del contractualismo moral liberal, la estructura del argumento que desarrollaremos y sus relaciones con otras teorías políticas y éticas contemporáneas.

a) Justificación moral y contrato.-

El contractualismo pretende justificar, ante un individuo racional autointeresado, la necesidad de establecer y mantener una determinada estructura normativa, y de respetar las obligaciones que sus reglas impongan[1]. El contractualismo moral afirma que esa estructura (o una parte de ella) corresponde a la normatividad moral, y que el cumplimiento no está condicionado por la imposición de límites externos (coacciones), sino que puede asimismo justificarse racionalmente. Dicho de otra forma, el contractualismo intenta responder afirmativamente a un problema clásico, cifrado en la pregunta ¿es racional ser moral? La originalidad de la respuesta contractualista a este interrogante tal vez destaque más en relación a alguno de sus puntos de

[1] Sobre la idea de "justificación" no podemos extendernos. Aceptamos de modo genérico la visión de Baier, expuesta en "Justification in Ethics" (en Pennock y Chapman [eds.], *Justification*, Nueva York, New York U.P., 1986, pp. 3-27. Cfr. especialmente p. 4), según la cual la justificación práctica de una institución, actividad o práctica social, ofrecida a sus miembros e, idealmente, a todos los miembros de la sociedad, consiste en mostrar que todos tienen razones adecuadas (quizá irresistibles) para querer que esa institución, actividad o práctica continúe, y que nadie tiene una razón adecuada para oponerse a ella. Esta comprensión de la justificación puede complementarse con el sentido que le da Richard Brandt en *A Theory of the Good and the Right* (Oxford, Clarendon, 1979; p. 183), según el cual una institución estaría justificada si (y en la medida en que) cumple *su* función. Justificar una institución consiste, por tanto, en comprobar si la misma cumple satisfactoriamente la función que se supone debe cumplir.

En nuestra opinión, una teoría contractual de la moralidad tiende a satisfacer ambos sentidos de la justificación: ofrece a los individuos argumentos para querer racionalmente la continuidad (bajo ciertas condiciones) de las instituciones examinadas, y a la vez contempla esas instituciones bajo la luz de su utilidad para el individuo o el grupo, de modo que sólo se consideran justificadas si cumplen la función asignada y ningún otro mecanismo más sencillo puede cumplir la misma función en el mismo grado. Ambos sentidos se complementan, porque sólo si una institución es funcionalmente justificable podrá vindicar el asentimiento racional de sus miembros.

referencia polémicos.

Un método relativamente frecuente para intentar demostrar la racionalidad de las restricciones morales ha consistido en el análisis de la argumentación práctica. Nagel, por ejemplo, creyó descubrir una exigencia racional sobre la acción en la estructura del argumento práctico. Esa exigencia consistiría en el altruismo (respecto a otros) y la prudencia (respecto a uno mismo)[2]. Puesto el altruismo como base de la moralidad, es racional ser moral porque es una necesidad de la razón el ser altruista.

Lo que intenta Gauthier, y el contractualismo en general, es derivar esa necesidad *no de la estructura del argumento práctico, sino de la estructura de la interacción*. El contractualista niega que el argumento práctico contenga o implique (ni en su substancia ni en su forma) principio o requerimiento moral alguno. Ello queda demostrado por el hecho de que, en casos de decisión paramétrica (como el ejemplificado por la operación del mercado perfectamente competitivo) no es posible establecer límite legítimo alguno a la libertad del agente racional. Por tanto, tampoco será legítimo establecer límites morales a partir de (o implicados por) la deliberación racional *per se*. Ahora bien, el estudio de la interacción muestra un carácter peculiar de la racionalidad como maximización: es colectivamente auto-frustrante[3]. La perplejidad ante esa característica (aparentemente inevitable) de la racionalidad es el punto de partida de la cooperación. En efecto, la cooperación es un modo de interacción que permite superar la contradicción de la racionalidad maximizadora. Pero la cooperación supone reglas, supone la limitación de la libertad individual (eventualmente supone incluso la renuncia individual a la maximización), supone una distribución pre-determinada del beneficio cooperativo, supone estabilidad a lo largo del tiempo, etc. Pues bien, todas estas ordenaciones carecen de una solución racional individual. No existen condiciones tales, ni reales ni hipotéticas, que puedan permitir a un sólo individuo —por más que

---

[2] Cfr. Nagel, T., *The Possibility of Altruism*, Oxford, Clarendon, 1970, p. 87.

[3] El término *self-defeating* ha sido traducido también como "auto-refutatoria" (Bayón Mohino). Podría usarse también "auto-cancelante" u otras expresiones análogas. Creo que el sentido es suficientemente claro. Como es sabido, proviene del análisis clásico de Derek Parfit en *Reasons and Persons* (Oxford, Clarendon, 1984), pp. 55 y ss.

represente a la Perfecta Racionalidad— decidir de una vez para siempre qué reglas o principios de la cooperación son más adecuados para una sociedad[4]. El único modo de poner las bases de la cooperación consiste en realizar un *pacto* tal que su cumplimiento sea tan racional para cada parte como racional es el hecho de suscribirlo. Y un pacto así sólo puede ser aquél en cuya formación o negociación los intereses de cada parte se encuentren directa, igual e individualmente representados y tenidos en cuenta, sin posibilidad de fraude, ni debilidad en el poder negociador de ninguna de las partes. Las condiciones del pacto quedarán fijadas por la Teoría de la Negociación Racional, que cumple estos requisitos generales. Ella nos proporciona un criterio para determinar los principios racionales de la cooperación que, en la medida en que reflejan imparcialmente los intereses de cada individuo, podrán exigir legítimamente ser universalmente aceptados, y se ajustarán a la definición de los principios morales.

El contractualista responde a la cuestión sobre la racionalidad de la moral retrocediendo hasta el origen de la misma. Acepta una serie de presunciones contra la moralidad para preguntarse, simplemente, qué es racional hacer desde el punto de vista de la maximización del auto-interés. Y en la búsqueda de una solución a ese problema general de la racionalidad y la acción, la moral aparece como un invitado imprevisto. La moral (un principio moral) resulta ser —en las aclaradoras palabras de Wolf citadas por Gauthier[5]— la solución de un juego de negociación que represente el contrato social. La solución a estos juegos es un principio moral porque, una vez que se ha elegido un principio sobre la base del propio auto-interés, entonces los negociadores (convertidos ya en miembros de la sociedad) están obligados por

---

[4] El desarrollo de esta idea, sobre la que, de todas formas, volveremos, puede verse especialmente en Gauthier, D., "Justice and Natural Endowment: Toward a Critique of Rawls's Ideological Framework", en Gauthier, D., *Moral Dealing*, Ithaca, Cornell U.P., 1990, pp. 150-170, y "The social Contract: Individual Decision or Collective Bargain", en Hooker, Leach and McClennen (eds.) *Foundations and Applications of Decision Theory, Vol. II*, Dordrecht, Reidel, 1978, pp. 47-67.

[5] Las referencias corresponden a Wolf, R.P., *Understanding Rawls. A Reconstruction and Critique of* a Theory of Justice, Princeton, Princeton U.P., 1977, pp. 13 y ss., citado por Gauthier en "Between Hobbes and Rawls", en Gauthier y Sugden (eds.), *Rationality, Justice and the Social Contract*, Ann Arbor, University of Michigan Press, 1993, pp. 24-39.

ese principio en todos los casos futuros, incluidos aquellos en que no resulta individualmente beneficioso tal sometimiento. Al aceptar el principio y, sobre todo, al obligarse por él, los individuos sufren una transformación moral.

El contractualismo moral pretende reconstruir racionalmente esa transformación, y con ello explicitar la base racional de la moralidad, de modo que pueda justificarse el comportamiento moral. Esa tarea de reconstrucción exige que la teoría genere "estrictamente como principios racionales para la elección, y por lo tanto sin introducir supuestos morales previos, restricciones sobre la persecución individual del interés o el beneficio que, siendo imparciales, satisfagan el concepto tradicional de moralidad"[6].

El contractualismo moral liberal descubre de modo clarificador la relación interna entre racionalidad y restricciones morales. Su idea nuclear es que nuestro acceso a esa relación sólo existe a través del despliegue heurístico de una hipotética negociación racional entre individuos pre-sociales. La teoría consiste en la "deducción" del marco normativo que surgiría de tal negociación; y en la medida en que tal deducción sea correcta (se siga de las premisas), la teoría puede considerarse válida. El convencimiento de que este despliegue teórico puede tener influencia sobre nuestra práctica concreta descansa en la certeza de que las premisas del argumento contractualista liberal son verosímiles; si efectivamente lo son, entonces la teoría será, además de válida, razonable y plausible como filosofía moral[7]. Lo característico del contractualismo es que la "deducción" a que hemos aludido no es una deducción lógica en sentido estricto, sino que tiene un componente que podemos denominar "histórico" o procedimental. Lo que media entre las premisas y la conclusión normativa no es un silogismo, sino la reconstrucción paso a paso del proceso que seguirían los individuos en el estado de naturaleza para salir del mismo. Tal como han sido establecidas las premisas, este proceso sólo puede tener una

[6] Gauthier, D., *MA*, p. 6.

[7] La distinción entre validez y plausibilidad del argumento contractualista procede de Jody S. Kraus. Cfr. su *The Limits of Hobbesian Contractarianism* (Nueva York, Cambridge U.P., 1993), p. 20.

dirección y resultado racionales, lo cual legitima la tarea reconstructiva en que la teoría consiste.

La originalidad del contractualismo moral liberal estriba en el uso de la negociación racional como mecanismo explicativo del contrato social. Es sabido que hay formas de contractualismo que confían en que el resultado del pacto social puede identificarse con una elección individual bajo ciertas condiciones privilegiadas. La idea de una negociación racional entre agentes auto-interesados no sólo es más coherente con el postulado individualista, sino que recoge de modo especialmente adecuado la intuición fundamental del contractualismo. Esto ha sido puesto de manifiesto por varios estudiosos de la contemporánea Filosofía Política norteamericana, entre ellos el francés Philippe Van Parijs[8]. La presencia del mecanismo de negociación racional distingue, por tanto, al contractualismo liberal cuyo argumento central vamos a resumir inmediatamente. Sin embargo, no es la única distinción que cabe efectuar, y antes de internarnos en las dificultades del argumento quizá sea conveniente situar geográficamente el contractualismo de Gauthier por referencia a otras teorías (contractualistas algunas de ellas) de la justicia. Estimamos que esta tarea de balizamiento intelectual ya ha sido anticipada en los capítulos anteriores, pero este es un buen momento para precisar con mayor rigor, en la medida de lo posible, lo que hasta ahora han sido referencias inconexas.

[8] En concreto, Van Parijs, escribe (en *¿Qué es una sociedad justa?*, Barcelona, Ariel, 1993 [trad. Juana A. Bignozzi], p. 210): "sólo en la tradición (propietarista) del beneficio mutuo el constructivismo puede ser contractualista en un sentido más estricto. En la tradición (solidarista) de la imparcialidad, en efecto, faltan al procedimiento hipotético varios elementos esenciales del contrato, en sentido económico del término: la no cooperación (el "estado de naturaleza"), es decir, la situación en ausencia de contrato, no es una base de referencia pertinente, ya sea porque (en la variante "posición original") el velo de la ignorancia transforme la negociación de los términos del contrato en un proceso de elección individual en condiciones de incertidumbre, o porque (en las otras variantes) la búsqueda de un acuerdo razonable es reemplazada por la búsqueda del interés personal. Sólo conservando estas reservas mentales puede ser esclarecedor hablar de "contractualismo" para designar no sólo el constructivismo propietarista de Gauthier, sino también el constructivismo solidarista de Rawls o Habermas".

b) El contrato moral y las teorías liberales de la justicia.-

Nadie va a confundir el contractualismo liberal con una teoría ética fenomenológica ni personalista, de forma que tales demarcaciones serían inútiles. Lo relevante es situar nuestra teoría en relación a otras teorías liberales de la justicia, algunas de las cuales dicen estar fundadas sobre los mismos principios que hemos señalado para el contractualismo moral liberal.

Brian Barry, en su conocida obra *Theories of Justice*, emplea varios criterios para clasificar las teorías liberales de la justicia según su estructura; nos serviremos de su conceptualización para intentar iluminar nuestra tarea de "situar" el contractualismo moral. Barry parte de la base de que todas las teorías liberales de la justicia pueden considerarse teorías "constructivistas" y "procedimentales"[9] y, por lo tanto, cabe establecer una diferencia relevante entre ellas según ciertos caracteres del procedimiento mediante el cual esperan construir o seleccionar principios de justicia. Algunos de los criterios elegidos son: la configuración de la posición inicial o línea-base desde la que se inicia el proceso de selección de los principios de la justicia, especialmente por referencia a la propiedad privada; la cantidad de conocimiento sobre la identidad de las partes en la posición inicial; el tipo de motivación que impulsa el proceso; el mecanismo de división del beneficio cooperativo, etc. La combinación de estos criterios permite una clasificación de, al menos, cuatro tipos de teorías según la configuración de la posición inicial y otros cuatro tipos de teorías según la estructura de la segunda fase del argumento (la negociación

[9] Esto significa, dicho muy a *grosso modo*, que todas estas teorías comparten la tesis que consiste en negar la existencia de una concepción substantiva de la justicia y sostener que, por tanto, no hay un criterio substantivo para elegir un principio de justicia social. Tal principio ha de ser seleccionado mediante la estrategia de diseñar un adecuado procedimiento (imparcial o "procesalmente justo") cuyo resultado definirá qué es lo justo, sin que pueda alegarse un criterio independiente del proceso mismo. el uso de esta justicia puramente procesal implica —en palabras de Rawls— que los principios de justicia mismos han de ser *construidos* por un proceso heurístico de deliberación. Como veremos, la diferencia entre distintas teorías constructivistas estriba en el diseño del proceso "imparcial" y la caracterización de las hipotéticas partes que han de tomar parte en él. Sobre estas distinciones y definiciones, cfr. Barry, B., *Theories of Justice*, Londres, Harvester-Wheatsheaf, 1989, pp. 264 y ss.

o decisión sobre los principios)[10]. Pero, en general, todas las teorías pueden inscribirse, según Barry, en uno de los dos grandes paradigmas: la justicia como beneficio mutuo o la justicia como imparcialidad.

De acuerdo con la taxonomía de Barry, la teoría moral de David Gauthier corresponde a una concepción de la justicia como beneficio mutuo, caracterizada por una posición inicial en que las partes tienen un conocimiento completo de su identidad e intereses personales y en la que la motivación esencial para iniciar un proceso encaminado a la cooperación social es el auto-interés. La línea-base que se toma como punto inicial de la negociación está determinada por el resultado de la acción maximizadora de cada agente, pero con ciertos límites que prohíben aprovecharse de la explotación de otros en el estado de naturaleza y, por último, el mecanismo de decisión sobre la distribución del beneficio cooperativo es una hipotética negociación ideal.

Podemos comparar, a modo de ejemplo, la teoría de Gauthier con algunos otros intentos constructivistas conocidos. Así, Barry situaría la teoría de Nash cercana a la de Gauthier (a salvo de las ambigüedades de la misma), sólo diferenciada por la admisión de la maximización sin restricciones en el estado de naturaleza, lo que modificaría la posición inicial de negociación y, por ende, el resultado. Otra teoría de la justicia como beneficio mutuo, la de Harsanyi, se distinguiría porque excluye la posibilidad de una información completa sobre la identidad personal en la situación original, lo que da lugar, como es sabido, a un principio utilitarista de justicia social adoptado, además, no mediante una negociación, sino mediante una decisión individual. Por supuesto, las concepciones de la justicia como imparcialidad —o "solidaristas", como las denomina Van Parijs— se distinguen todas ellas porque descartan la idea de una negociación como mecanismo de elección de los principios. Unas (como la de Braithwaite) adoptarían como línea-base el resultado de la libre interacción natural; otras (como la de Rawls) limitan el punto de partida con fuertes restricciones igualitaristas. Quizá lo más característico de estas teorías, frente a las representadas por Gauthier o Harsanyi, es que han de suponer que existe una motivación distinta del auto-interés para iniciar el proceso que

[10] Cfr. Barry, *op. cit.*, pp. 293 y ss. y 320-322.

conduce hacia una sociedad justa. Barry habla de un "deseo de alcanzar un acuerdo sobre términos razonables". En estas teorías, la tendencia a la sociedad (y a establecer una sociedad aceptablemente justa) es un supuesto inicial. En términos sencillos la diferencia sería esta: las partes en la posición inicial de Rawls se dicen "hagamos una sociedad justa; y, para ello, decidamos qué principio de justicia distributiva ha de regir sus instituciones"; mientras las partes del contrato de Gauthier se dicen "veamos qué es más beneficioso; resulta que lo más beneficioso es cooperar, así que debemos ponernos de acuerdo sobre un principio de distribución que haga posible la cooperación". Para el primer tipo de razonamiento, la motivación es un deseo de cooperar y el problema, decidir qué principio hará más justa (imparcial) esa cooperación. Para el segundo, la motivación es el beneficio individual, y el problema es el acuerdo mismo que, al garantizar la cooperación, permite maximizar el beneficio.

Estas indicaciones, basadas en la taxonomía de Barry, pueden situar con alguna (limitada, por cierto) claridad el contractualismo moral, pero al coste de encasillar las teorías con una rigidez completamente artificial. Tal vez las teorías más "puras", esto es, las que corresponden de modo estricto a uno de los extremos del espectro conceptual empleado como criterio de clasificación, admitan esta rigidez (aunque es improbable; de hecho, Barry ha de interpretar un tanto libremente el pensamiento de varios autores para conseguir que sus respectivas teorías encajen en sus esquemas), pero éste no es el caso de la moral por acuerdo de Gauthier. La originalidad del enfoque del canadiense se pondrá de manifiesto precisamente en la dificultad de clasificar adecuadamente la teoría.

Esta dificultad aparece meridianamente en el estudio de Van Parijs (aunque para el francés, consciente del novedoso intento de Gauthier, no representa ningún problema). Entre las clasificaciones que encontramos en *¿Qué es una sociedad justa?* destacan un par de distinciones muy adecuadas para hacer ver la originalidad del enfoque de Gauthier. La primera distingue entre teorías *propietaristas* y teorías *solidaristas*[11]; la segunda, entre teorías

[11] Cfr. Van Parijs, P., *op. cit.*, p. 200 y ss.

retrospectivas y *prospectivas*. El liberalismo propietarista define una sociedad justa como aquella que no permite que a un individuo se le arrebate lo que le corresponde según unas reglas pre-definidas. Esta concepción de la justicia (representada por los libertarios radicales, como Rothbard) concede absoluta preeminencia a la apropiación individual, sin límite alguno, que abarca el derecho a las propias capacidades naturales, a las creaciones del trabajo directo y a lo adquirido originariamente como primer ocupante, así como a lo adquirido mediante intercambio voluntario. Cualquier interferencia en las propiedades y derechos individuales así obtenidos se considera ilegítima. Para el liberalismo "solidarista", por el contrario, la sociedad justa no sólo debe tratar a los individuos con igual respeto, sino también con igual cuidado, atención o asistencia. El problema de la justicia no es el respeto escrupuloso de las propiedades individuales, sino la *distribución* de una cierta variable (que en general podemos identificar con "el bienestar"). Este problema puede solucionarse de acuerdo a principios estrictamente distributivos (como los principios de la justicia de Rawls) o hacerse depender de principios agregativos, como el principio de utilidad clásico.

Por otro lado, Nozick distinguió entre teorías retrospectivas, que consideran la justicia desde el punto de vista de la genealogía de los títulos o derechos (es justa la distribución basada en derechos legítimos y en transacciones voluntarias), y teorías prospectivas: aquellas que consideran la justicia desde el punto de vista de los fines de la acción. En esta doble conceptualización sirve para situar a la mayoría de teóricos liberales contemporáneos, pero veamos cómo podríamos aplicarla a Gauthier: Frente al liberalismo netamente solidarista de Rawls, Gauthier se definiría como propietarista, en el sentido de que la justicia social se expresa primeramente en el reconocimiento del derecho que cada individuo tiene a "lo que trae a la mesa de negociación", es decir, a sus propiedades y derechos pre-contractuales. Éstos representan un límite para los posibles principios distributivos, pues nadie aceptaría una regla cuyo resultado fuese un coste neto, para él, por el hecho de entrar en la sociedad. Sin embargo, frente al propietarismo de los libertarios, incluso frente al propietarismo limitado de Locke y Nozick, las restricciones que Gauthier impone a la configuración de la posición inicial de negociación implican, de

hecho, que la estructura de la propiedad privada que ha de ser respetada y defendida por el principio de justicia está también, de algún modo, co-determinada por la negociación (ya que, como veremos, las exigencias de una negociación racional se extienden a la "razonabilidad" de la posición inicial). Frente al clásico enfoque retrospectivo de los propietaristas, el enfoque de Gauthier es predominantemente prospectivo. Aunque se acepta que las partes de la negociación son individuos con su historia, sus capacidades naturales y sus derechos adquiridos, ello es sólo con un carácter provisional, porque en realidad es la lógica de la negociación misma la que determina qué configuración pre-contractual de derechos resulta aceptable como base para negociar. De este modo, el resultado del proceso, aunque tiene en cuenta las capacidades naturales y los intereses de cada agente —de modo que nadie tiene base para considerarse perjudicado por la entrada en la sociedad— es eminentemente prospectivo, pues intenta instaurar un principio distributivo finalmente proporcional.

Estas breves notas no pueden sino señalar aproximadamente la localización ideológica, según los estudiosos del tema, de la teoría que venimos analizando. Pero tampoco pretendíamos otra cosa en este momento. Intentaremos en lo que sigue desarrollar en detalle nuestra comprensión del contractualismo moral liberal de modo que el difícil equilibrio entre propiedad y solidaridad, entre auto-interés y moralidad, que se habrá filtrado en los párrafos anteriores aparezca, como lo hace ante nuestros ojos, con mayor evidencia.

c) Estructura de la teoría del contrato moral.-

En el capítulo inicial de *MA*, Gauthier reduce a cuatro las ideas centrales de su teoría: la salvaguardia lockeana que prohíbe el beneficio obtenido de la explotación de otros en el estado de naturaleza; la zona exenta de moralidad ejemplificada por el mercado perfectamente competitivo; el principio distribu-

tivo de la concesión relativa *minimax* y la disposición a la maximización restringida[12]. Ya hemos comentado que el papel central que Gauthier asignaba en 1986 al mercado perfecto en su teoría ha de ser matizado. Ello nos deja con tres concepciones centrales, de las cuales una, el principio de la concesión relativa *minimax*, forma parte (o es consecuencia) de un todo mayor y complejo, a saber, la teoría de la negociación racional. Así pues, la negociación, la salvaguardia y la maximización restringida son los pivotes de la teoría en su parte más abstracta y técnica. Otros elementos relevantes —pero relativamente alejados del núcleo teórico de la obra— serán la idea de un "punto arquimédico", que sirve básicamente, en nuestra opinión, para contrastar el resultado de la moral por acuerdo con las conclusiones de la *Teoría de la justicia* de Rawls; y la concepción del *individuo liberal*, que puede tomarse como verdadero fruto del esfuerzo teórico de Gauthier, al presentar la imagen acabada de las relaciones entre las personas e instituciones en una sociedad liberal tal como cabe concebirla sobre la base de una moral contractual y, sobre todo, establecer el papel y estatuto ontológico de dicha moral.

Las tres ideas centrales que hemos distinguido configuran, tomadas en conjunto, un argumento contractualista. Como sabemos, este tipo de teorías adoptan la forma de una historia desarrollada en el tiempo: se describe una situación inicial; se explicita el tipo de lógica que, afectando a todas las personas que se encuentran en esa situación, conduce a la conclusión de que la cooperación (esto es, la sociedad) proporcionará ventajas mutuas; se contempla cómo se desarrollaría, paso a paso, una negociación o diálogo y cuál sería, finalmente, su resultado; y se comprueba después qué respuesta darían los ciudadanos a las obligaciones derivadas de su compromiso. Si esta respuesta consiste en cumplir lo pactado, la teoría del contrato ya ha alcanzado su objetivo justificador; si, por el contrario, la respuesta individual más probable consiste en el incumplimiento, entonces hay que dar un paso más e implementar medidas que impidieran la quiebra del pacto.

En definitiva, el modo expositivo tradicional de las teorías del contrato,

[12] Cfr. *MA*, p. 16.

en forma de sucesión, hace olvidar muchas veces que toda la teoría no es más que una reconstrucción racional que carece de sentido si no es comprendida globalmente. Esto se aplica a cualquier teoría del contrato, incluyendo el contractualismo clásico, pero es más evidente en el caso de Gauthier, debido a la complejidad del argumento.

La idea unitaria que guía el despliegue sucesivo de las nociones centrales y su discusión es que hay razones auto-interesadas para someterse voluntariamente a aquel principio de la justicia que resultase de una negociación ideal, perfectamente racional, entre agentes situados en una posición inicial no-coactiva, esto es, en la que cada cual tuviera, como "equipaje natural", todas sus capacidades y los bienes o expectativas adquiridos mediante su uso, siempre que en esa adquisición no hubiese ocasionado una pérdida efectiva en los bienes, expectativas o capacidades de otro futuro miembro de la sociedad. A su vez, este límite en lo que cada cual trae a la mesa de negociación como "dotación natural" no es caprichoso, sino que representa el requisito mínimo para que el resultado de la negociación sea estable (es decir, que sea racional para todos los miembros de la sociedad cumplirlo). Por último, todo el engranaje de la negociación (que incluye la determinación de la posición inicial razonable) se pone en marcha sólo si se supone que existe una capacidad individual de "racionalidad crítica", esto es, de cuestionar los propios criterios de racionalidad radicalmente, hasta el punto de variarlos cuando es estrictamente racional —según los criterios actuales— hacerlo. Esta capacidad será fundamental para poder justificar el cumplimiento *ex post* de lo que se admitió como racional *ex ante*.

De este modo, las tres ideas nucleares de la moral por acuerdo son lógicamente simultáneas, aunque el despliegue teórico sea necesariamente sucesivo. El hablar primero de la negociación quiere indicar que las otras dos ideas, la salvaguardia y la maximización restringida, son, en cierta medida, *condiciones* para que la negociación tenga sentido. Pero son condiciones que se hacen presentes sólo *en* la negociación (o, al menos, *a la vista de* la negociación), por lo que sería incorrecto pensarlas como condiciones *previas*.

La incomprensión de la aludida simultaneidad lógica de los elementos de la teoría ha dado lugar a críticas cuya futilidad debe quedar de manifiesto tras

nuestro análisis. Se ha querido interpretar, por ejemplo, la salvaguardia lockeana como un límite moral a la adquisición de propiedad en el estado de naturaleza. Pero tal interpretación, sin referencia a la negociación, la cooperación y el pacto, es completamente equívoca, y lleva a confundir el enfoque de Gauthier con un libertarismo mal fundamentado. También se ha criticado la teoría de la negociación racional desgajándola de su contexto. Habrá que admitir que, como método matemático, el de Gauthier ha resultado no ser generalizable, pero es cuando menos dudoso que ello tenga consecuencias decisivas para la teoría como un todo. Tal vez el concepto de maximización restringida admite un tratamiento más independiente. De hecho, es un concepto que, tanto antes como después de *MA*, Gauthier ha procurado desarrollar independientemente, aunque siempre en relación con la posibilidad de justificar una moral por acuerdo[13].

La dificultad para alcanzar una comprensión global de la teoría contractual puede entenderse si se tiene en cuenta que se trata de un método heurístico que aparece en el panorama filosófico-político moderno precisamente ante la escasa verosimilitud de los argumentos tradicionales, formalmente deductivos, cuya premisa mayor era la autoridad divina o la ley natural. Por su propia naturaleza, el argumento contractualista no coincide formalmente con lo que intenta significar. Todo él es una especie de metáfora: una historia inventada que quiere "representar" un nexo lógico: el que se da entre el interés individual y la justicia, entre razón y moralidad. La idea del contrato liberal aparece como tabla de salvación para escapar del escepticismo, pero a costa de expresar perifrástica, y tal vez inadecuadamente, la relación entre moral y

[13] El primer ensayo maduro y completo sobre la racionalidad restringida fue "Reason and Maximization" (*Canadian Journal of Philosophy*, 4 (1975), pp. 411-433). La misma idea es empleada, posteriormente, en otros artículos, y re-formulada con un gran aparato matemático (y con el nombre de "cooperación condicional") en "The Incompleat Egoist" (en Sterling y McMurrin (eds.), *The Tanner Lectures on Human Values, vol. 5*, Salt Lake, University of Utah Press, 1984, pp. 67-119). La siguiente formulación del concepto es la de *MA*, que debe más al primer artículo que al segundo. Tras *MA*, Gauthier ha seguido explorando la posibilidad de un compromiso racional que significase un giro hacia la restricción de la maximización. Su "Assure and Threaten" (*Ethics*, 104 [julio 1994], pp. 690-721) resume su progreso en este sentido, orientado en los últimos tiempos por los estudios de Edward F. McClennen en torno a la decisión dinámica, expuestos en *Rationality and Dynamic Choice: Foundational Explorations* (Cambridge, Cambridge U.P., 1990).

racionalidad. El contrato social *capta* —esta es la convicción que compartimos con los restauradores contemporáneos de la teoría— la relación entre racionalidad individual y moralidad, mas el modo procedimental de la exposición difumina parte de la intuición básica que intenta transmitir.

En este contexto, no es reprochable interpretar la obra de Gauthier como "solamente" un intento de explicar técnicamente la cooperación, o de aportar una solución posible para el problema colectivo de la provisión de bienes públicos, o de defender una concreta posición sobre los derechos naturales; aunque todas ellas son interpretaciones incorrectas de puro incompletas y sesgadas.

En la medida de nuestras posibilidades, intentaremos ofrecer una lectura global de la moral por acuerdo. Esta lectura sólo puede consistir en mantener siempre la conciencia de que cada parte del argumento es una pieza de una unidad mayor que funciona como un todo, y que ha de ser criticada o admitida como un todo. Por lo demás, la exposición habrá de quedar dividida inevitablemente —como sucede en el *MA*—, tanto por la naturaleza del argumento, como por las dificultades particulares que cada problema plantea.

Seguiremos el orden expositivo de Gauthier. No es un orden arbitrario, como hemos comentado antes, pero tampoco necesario. Es decir, en cuanto las piezas de la teoría sólo funcionan una vez que se engranan en un sólo mecanismo teórico, carece de importancia cuál de ellas se expone en primer lugar; ahora bien, el orden elegido intenta facilitar la comprensión. Especialmente hay que destacar que el hecho de hablar en último lugar de la posición inicial de negociación y sus límites obedece al declarado propósito de hacer ver que la configuración de la misma depende completamente de la posibilidad de la cooperación.

Iniciaremos el comentario caracterizando la interacción en el estado de naturaleza. Esta parte ha sido anticipada en los comentarios previos sobre el postulado individualista, la racionalidad y la zona exenta de moralidad, por lo que nos liberaremos de una exposición demasiado tediosa. Después seguiremos con la lógica del contrato, pasando al análisis de la negociación como mecanismo para salvar el problema de la cooperación. Una vez concluido el pacto, la atención se vuelve hacia la racionalidad de cumplir lo pactado, caballo

de batalla del contractualismo desde sus orígenes. Gauthier plantea este problema en los términos clásicos, recordando la objeción del *Foole* en el *Leviathan*, pero intenta solucionarlo de modo completamente novedoso (aunque muy cuestionable), mediante el expediente de la maximización restringida. Por último, y con el propósito antedicho, se estudiará la configuración de la posición inicial de negociación. Por último, retornaremos al tipo de discusión al que hemos apuntado en estos últimos párrafos, es decir, intentaremos releer globalmente la teoría para, expuestas las críticas a la misma, ofrecer su sentido como filosofía moral.

2. *La interacción natural y el "Dilema del Prisionero"*

a) La irracionalidad de elegir el egoísmo.-

Este epígrafe es el título de un breve artículo de Gauthier escrito como respuesta a una crítica vertida contra "Reason and Maximization" por Larry Eshelman[14]. Lo empleamos porque, junto a otras incursiones de Gauthier en el problema del egoísmo como principio ético[15], nos presta un lenguaje inigualable para formular los dilemas de la interacción natural. La segunda fuente conceptual que emplearemos proviene de nuestro tratamiento del mercado perfectamente competitivo, en el punto 4 del capítulo segundo. Allí nos oponíamos a la lectura que interpreta el mercado como el equivalente del "estado de naturaleza" en la teoría de Gauthier, pero también admitíamos su función como modelo de interacción perfectamente racional por referencia al cual caracterizar la acción estratégica no-cooperativa propia de la situación pre-contractual.

Recordaremos, por tanto, que el mercado perfectamente competitivo incluye una serie de postulados contra-fácticos cuya remoción nos aboca a los conocidos fallos del mercado real. Estos fallos, bien conocidos en la práctica económica ultraliberal, no dependen de la falta de habilidad, destreza o racionalidad de los agentes, sino de ciertas condiciones dadas, como puede ser

---

[14] "The Irrationality of Choosing Egoism: A Reply to Eshelman", *Canadian Journal of Philosophy*, 10 (1990), pp. 179-187. Podemos anotar la curiosidad de que, si bien Gauthier se deshace de la crítica contra su idea de la maximización restringida, descubre que el texto de Eshelman apunta inintencionadamente una falla insospechada en su formulación. Para solventar esta falla, Gauthier hubo de integrar la teoría de la negociación racional en su visión de la moralidad como un "compromiso racional para abandonar el egoísmo".

[15] Gauthier, D., "The Impossibility of Rational Egoism", *The Journal of Philosophy*, vol. LXXI, nº 14 (agosto, 1974), pp. 439-456. En este texto Gauthier despliega una interesante —y, desde el punto de vista de la Teoría de la Decisión Racional, definitiva— crítica contra el egoísmo ético.

la existencia de "bienes libres", el coste de la información, etc. Ya mencionamos que los fallos del mercado pueden reducirse a la aparición de externalidades y parásitos, junto con la infra-producción de bienes públicos. Dado que estos problemas de la interacción de mercado no suponen carencia alguna en la racionalidad de las partes, se pueden tratar como un "problema técnico" derivado de condiciones que, en otro contexto, han sido denominadas "circunstancias de la justicia"[16], en el sentido de que son las circunstancias en que se hace necesario establecer medidas coordinadas —habría que decir, mejor, cooperativas— de producción y distribución de bienes. Tales medidas no serían necesarias en una situación como la descrita por el mercado ideal perfectamente competitivo (donde reinaría lo que Buchanan denomina "anarquía del mercado"[17]). Justamente eso es lo que se quería indicar al calificarla como una zona "exenta de moralidad": una zona donde la justicia estaría injustificada desde el punto de vista individual, pues —como cualquier intervención supondría una modificación del resultado *paretiano* de mercado— no habría un criterio de racionalidad colectiva suficientemente pregnante que oponer a la maximización individual.

Así pues, los fallos del mercado evidencian que nos hallamos, en general, bajo las condiciones de la justicia. El análisis del resultado de la interacción de mercado nos ofrece el siguiente panorama: cada agente persigue maximizar su interés; sin embargo, todos terminan en una situación netamente inferior (en términos de utilidad) a aquella que podrían alcanzar si, ocasionalmente, renunciasen a su interés privado para contribuir a la producción de bienes públicos. Es decir, se establece un contexto en que la justicia resultaría *útil*.

No es sorprendente que esta formulación se avecine a la derivada del Dilema del Prisionero. También el siguiente paso en el razonamiento maximizador es común en ambos casos: una vez establecidas las bases de la

---

[16] Cfr. David Hume, *Treatise*, libro III, parte II, cap. II. Hume habla de una "moderada escasez" y cierta tendencia egoísta natural. Esas son las circunstancias que harían tanto posible como necesaria la justicia.

[17] Cfr. Buchanan, J., *The Limits of Liberty*, Chicago, University of Chicago Press, 1975, pp. 18-19.

justicia (mediante un sistema de recaudación y administración de impuestos, por ejemplo) sigue siendo individualmente más racional el comportamiento maximizador (defraudar) que el cooperativo (cumplir lo pactado, o lo reglamentado). Ello es así porque, supuesta la cooperación de los demás, el beneficio (la producción de bienes públicos) está garantizado, con lo que es individualmente más ventajoso no contribuir; y en el caso de que el fraude esté tan extendido que la producción de bienes públicos sea imposible, entonces también será más beneficioso abstenerse de contribuir, para no incurrir en un coste sin contrapartida alguna. Este tipo de razonamiento parece bloquear el camino hacia la cooperación, por muy evidente que aparezca la utilidad de ésta.

La misma idea intuitiva que acabamos de exponer empleando el lenguaje del mercado y del llamado "dilema del contribuyente" se ha formulado técnicamente a través de la Teoría de Juegos. El Dilema del Prisionero muestra insuperablemente esta paradoja, como es sabido. Nos detendremos lo imprescindible para recordar la estructura de este dilema.

Dos sujetos *A* y *B* han sido detenidos acusados de un grave delito. Cada uno de ellos, encerrado en una celda de aislamiento, es interrogado e invitado a confesar, con la promesa de que si denuncia al otro detenido, saldrá en libertad (si no hay pruebas en su contra) o se le reducirá la pena. Si no confiesa, será de todos modos acusado de un delito menor por el que se le castigará con dos años de cárcel. Por otro lado, cada preso sabe que si el otro confiesa, no tendrá forma de evitar una condena muy larga, ya que ambos cometieron el delito grave. La conocida matriz es la siguiente (los números representan años de cárcel):

		B	
		confiesa	no confiesa
A	confiesa	10, 10	libre, 12
	no confiesa	12, libre	2, 2

Y, ordenando los resultados de más a menos preferido (los números representan el orden de preferencia de cada resultado):

		B	
		confiesa	no confiesa
A	confiesa	3°, 3°	1°, 4°
	no confiesa	4°, 1°	2°, 2°

En ambas matrices se aprecia que, desde el punto de vista de cada jugador, es prudente confesar, haga lo que haga el otro. Imaginemos el razonamiento de *A*: "si *B* confiesa, los posibles resultados serán 10 años de cárcel si confieso, 12 si callo, luego confesaré; si *B* no confiesa, mis pagos son la libertad si confieso o dos años de cárcel si callo, luego confesaré". El mismo razonamiento llevará a *B* a confesar, con lo que el punto de equilibrio de este juego es el producto de la confesión de ambos: diez años de cárcel para cada jugador, o sea, el tercer resultado, por orden de preferencia, para ambos. ¿No podrían ambos jugadores hacer algo mejor? Parece que sí. Salta a la vista que el producto de sus respectivos silencios beneficia a ambos en relación con el resultado de equilibrio o solución del juego. Si se pudiese alcanzar ese resultado (el segundo mejor para cada uno), ¿no sería racional hacerlo?

Dado que, por definición, ambos jugadores son agentes perfectamente racionales, los dos contemplarán lo absurdo que resulta pasar diez años en la cárcel teniendo a su alcance la posibilidad de una condena de sólo dos años. Si cada uno de ellos tuviera la garantía de que el otro va a callar, entonces tal vez callaría también y se lograría el resultado óptimo. Sin embargo ¿ocurriría realmente esto? Imaginemos que se permite a los detenidos hablar entre sí una vez conocida la oferta de los interrogadores. A la vista de la matriz que expresa los resultados posibles del juego, se pondrían de acuerdo en no confesar. Desde luego que cada uno prefiere la situación en que él mismo confiesa y el otro calla, pero entre maximizadores racionales no es de esperar semejante sacrificio por parte del otro; tal decisión ejemplificaría una conducta altruista que es

justamente la que hemos desterrado de este análisis al suponer que ambos son maximizadores racionales. Acordarían, por tanto, el mutuo silencio y, después, cada preso vuelve a su celda y se enfrenta al siguiente problema:

		B	
		defrauda	cumple lo pactado
A	defrauda	10, 10	libre, 12
	cumple lo pactado	12, libre	2, 2

Es decir, cada agente está exactamente en la misma situación que antes del pacto. Sigue siendo individualmente más beneficioso confesar (incumpliendo el pacto) que atenerse a lo pactado y arriesgarse al incumplimiento de la otra parte.

Este juego clásico no necesita más comentario: el Dilema del Prisionero hace patente una contradicción insoluble de la racionalidad maximizadora[18]. Dicho con las palabras asimismo clásicas de Parfit, "con frecuencia es verdad que si cada uno, en vez de ninguno, hace lo que es mejor para él, esto será peor para todos"[19]. La misma verdad se puede formular diciendo que la maximización individual ocasionalmente imposibilita la optimización colectiva. Y en la perplejidad ante este hecho hunde sus raíces el contractualismo:

> "La idea fundamental del contractualismo es que, bajo circunstancias que prevalecen ampliamente, el resultado de una conducta conforme con los estándares convencionales de la racionalidad

[18] El carácter insoluble de esa contradicción es enfatizado por algunos teóricos de juegos, para quienes cualquier intento de resolver el dilema opera, bien desvirtuando el juego, bien cometiendo alguna falacia. En este sentido, puede verse el análisis de Ken Binmore, "Bargaining and Morality", en Gauthier y Sugden, *Rationality, Justice and the Social Contract*, Ann Arbor, The University of Michigan Press, pp. 131-156; sección 2, pp. 136-140.

[19] Parfit, D., "Prudencia, moralidad y el Dilema del Prisionero", *Diálogo Filosófico*, 11 (1989), pp. 4-30; p. 7.

económica es incompatible con el logro de un óptimo."[20]

Gauthier aceptó el reto planteado por el carácter dilemático de la racionalidad maximizadora, e intentó demostrar que el egoísmo puro es lógicamente inconsistente como principio ético (o prudencial)[21] y, por otro lado, que entendido como maximización sin restricciones es contradictorio[22]. Con ello, se pretendía explicar la causa lógica del dilema de la racionalidad. A la vez, la conciencia de que se trata de un problema rigurosamente insoluble en los términos estrictos de la maximización individual, debe conducir a la reflexión sobre algún modo de interacción capaz de superar el dilema manteniéndose de acuerdo con los objetivos de la teoría normativa de la racionalidad, es decir, la maximización[23].

Como el mismo Dilema del Prisionero muestra, la *cooperación* (en forma de pacto) podría establecer un nuevo modo de interacción que permitiese alcanzar un óptimo, pero sólo si fuera posible que los agentes actuaran contra sus intereses directos en el momento de la elección. Ya que, mientras el paradigma individualmente maximizador de la racionalidad no se abandone, la cooperación tiende a frustrarse, pues sigue siendo individualmente más ventajoso no cooperar.

---

[20] Gauthier, D., "Economic Rationality and Moral Constraints", en *Midwest Studies in Philosophy*, III, 1978, pp. 75-96; p. 90.

[21] Cfr. "The Impossibility of Rational Egoism", cit.

[22] Cfr. "The Incompleat Egoist", en Gauthier, D., *Moral Dealing*, cit., pp. 234-273.

[23] Podría cuestionarse que el nuevo modo de interacción siguiera apegado al objetivo de maximizar la utilidad individual. Se podría argumentar, por ejemplo, que el fracaso de la racionalidad maximizadora muestra precisamente que se trata de una forma "defectuosa" de racionalidad, y que debe dejar paso a una racionalidad "altruista" (o universalista en general, como prefiere decir Gauthier). Según este paradigma alternativo, el objetivo de la racionalidad no es la maximización de la utilidad individual sino la satisfacción de algún criterio o medida pretendidamente objetiva. Creo que el empleo del paradigma de la racionalidad como maximización individual quedó suficientemente justificado en su momento; pero si ante el fracaso de la maximización individual que acabamos de exponer tan crudamente surgieran nuevas dudas, remitimos a un reciente artículo de Gauthier, "Public Reason", *Social Philosophy and Policy*, vol. 12: 1 (1995), pp. 19-42 (cfr. esp. pp. 19-22 y 42), donde se defiende que la única fuente de normatividad superviviente al "desencantamiento" normativo del mundo moderno ha sido la racionalidad individual.

b) La racionalidad de la cooperación.-

Lo dicho hasta aquí muestra, según el contractualismo moral liberal, el carácter del tipo de interacción que podríamos encontrar en un hipotético "estado de naturaleza". El hecho de emplear el Dilema del Prisionero como formalización de la interacción natural es una constante del neo-contractualismo de raíz hobbesiana[24], desde el estudio clásico de Gauthier[25]. Según este análisis, el estado de "guerra de todos contra todos" no es sino expresión intuitiva del dilema de la racionalidad económica: la competencia, rasgo esencial de la interacción estratégica, deja a todos los agentes en un estado paretianamente inferior a aquél que podrían alcanzar mediante la cooperación.

La cooperación, por tanto, se plantea ante la razón de cada agente como un *medio* para alcanzar su fin: la maximización de utilidad *individual*. La cooperación equivaldría a los "convenientes artículos para la paz" que la razón sugiere a cada hombre hobbesiano. Pero un paso más de la mano de Hobbes nos introduce en un círculo sin salida. Porque los hombres, conocedores de que no pueden escapar a su egoísmo natural, se auto-impondrían (por razones egoístas) un sistema de coacciones para garantizar el cumplimiento del pacto y, por ende, el resultado cooperativo. Sin embargo, este resultado, con ser mejor que el producto de la interacción natural, sigue sin ser óptimo. Aún hay un esquema de interacción paretianamente superior al Leviatán hobbesiano, y es el esquema en que se actuase cooperativamente *sin soportar el coste de la coacción*. Tal vez dicho esquema sea una contradicción en los términos, o inalcanzable en la práctica; pero, en todo caso, el camino que hacia él apuntase sería el que deberían esforzarse en recorrer los individuos auto-interesados, pues les promete el logro simultáneo de la optimalidad y la maximización.

Este es el camino del contractualismo moral. A través de él, David

---

[24] Cfr. Kraus, Jody S., *The Limits of Hobbesian Contractarianism*, Nueva York, Cambridge U. P., 1993, p. 11 y ss. Según Kraus, Hampton analiza el estado de naturaleza como un Dilema del Prisionero (DP) iterativo; Kavka, como un cuasi-DP; y Gauthier, como un DP único.

[25] Cfr. Gauthier, D., *The Logic of Leviathan*, Oxford, Clarendon, 1969, pp. 85 y ss. Que sepamos, se trata del primer análisis formal del contractualismo hobbesiano que emplea el DP para explicar la estructura de la interacción en el estado de naturaleza.

Gauthier pretende establecer la racionalidad de someterse voluntariamente a restricciones internas que, jugando el papel del soberano del Leviatán, permitan la cooperación. Así se inicia el programa de un contrato moral, frente al contenido político característico de la tradición.

Quisiéramos ahora introducir algunas de las dificultades iniciales que planteó el intento contractualista de demostrar que es racional ser moral. Estos comentarios nos servirán asimismo para indicar el proceso mediante el cual se ha de justificar la racionalidad de la cooperación.

El primer escollo que se plantea a la cooperación es la duda sobre el carácter auto-refutatorio de la teoría de la racionalidad como maximización del auto-interés. En efecto, el resultado del Dilema del Prisionero, aunque sub-óptimo, es individualmente maximizador, dadas las circunstancias. La acción de confesar maximiza efectivamente la utilidad esperada de cada agente. Así pues, la teoría cumple su objetivo. En palabras de Parfit, la teoría no es *individualmente* auto-refutatoria. Al contrario, lo que resulta contrario al objetivo maximizador propuesto por la teoría es el cumplir con lo que la cooperación demanda. Este es el motivo de que, incluso tras el pacto, no existan razones suficientes para cambiar de estrategia[26].

Parece, por tanto, y así es corroborado por el análisis del Dilema del Prisionero, que no es racional cooperar. El hecho de que la teoría de la racionalidad sea *colectivamente* auto-refutatoria —esto es, que cuando *todos* (por separado) buscan la maximización individual, entonces *todos* acaban en una posición que no es un *maximum*— tal vez justifique la imposición de restricciones sobre la libre maximización individual, pero estas restricciones

[26] Este tipo de razonamiento acude inmediatamente tras el primer análisis del DP. Así le sucedió a Gauthier, quien escribe, refiriéndose a sus primeras reflexiones sobre el famoso juego: "Al principio vi oscuramente. Vi en el DP una clara representación del conflicto entre razones de interés y razones morales o cooperativas, pero ninguna me parecía predominante. Vi un conflicto entre dos formas de racionalidad —una individual y prudencial, otra colectiva y moral. Y dije que "el individuo que necesita una razón para ser moral que no sea ella misma una razón moral, no puede tenerla... Porque es más que evidentemente paradójico suponer que cálculos de beneficios puedan alguna vez justificar la aceptación de una desventaja real". Yo estaba equivocado. Y es aquella paradoja, supuestamente genuina, lo que quiero refutar, para mostrar que se puede tener, y de hecho se tiene, una razón no-moral para ser moral, una razón que debe ser reconocida incluso por el egoísta" ("The Incompleat Egoist", cit., p. 255).

nunca serán vistas como racionales por el agente. Desde su punto de vista, la cooperación exige de él un tipo de comportamiento irracional (sobre esto volveremos más abajo, al comentar la racionalidad de cumplir los acuerdos). El contractualista individualista ha de afrontar la tarea de "convencer" al agente racional de que, cuando menos, merece la pena considerar la posibilidad de la cooperación.

Esta tarea se realiza básicamente mediante lo que podemos llamar "interiorización" de la optimización como un objetivo de la racionalidad. Un óptimo de Pareto puede ser definido como aquel estado de cosas en que el incremento de utilidad de un agente implica necesariamente una disminución en la utilidad de otro. Un estado óptimo supone una distribución de la utilidad tal que cualquier otra distribución alternativa supone un coste para algún agente. La optimalidad es un concepto que pone en relación a los agentes. Es un objetivo colectivo, que escapa al horizonte individual.

Sin embargo, en la medida en que es un objetivo racional, puede ser interpretado en términos de maximización individual[27]. Porque si el resultado de equilibrio en una situación de decisión estratégica es sub-óptimo, tal como acontece en el Dilema del Prisionero, *ningún* agente racional puede objetar un movimiento hacia el resultado óptimo ya que, a diferencia de lo que ocurre cuando se trata de establecer una nueva distribución en una situación ya óptima (cambiar de un óptimo a otro), tal movimiento significará un incremento de utilidad en uno o más agentes (en el caso del Dilema del Prisionero, *todos* se benefician) *sin coste alguno para nadie*. La optimalidad significa, literalmente, una utilidad igual o mayor para todos, respecto a la posición de equilibrio natural[28]. La situación óptima es, según esto, deseable desde el punto de vista

---

[27] Aunque esto no es cosa sencilla, como reconoce Gauthier: "Mostrar que el egoísmo es auto-refutatorio no es fácil. Como veremos, no basta con mostrar que los egoístas, al maximizar el valor relativo-al-agente, no obtienen tanto beneficio *colectivamente* como podrían. No basta con mostrar, en las palabras de Baier, que "la persecución del auto-interés por parte de todos es perjudicial para todos". Más bien debemos mostrar que la persecución del auto-interés por parte de cada persona es perjudicial para él, que cada uno deja de obtener *individualmente* tanto beneficio como podría." (*ibid.*).

[28] Estas ideas conectan con la creencia de que la cooperación es capaz de *producir* utilidad. Es decir, que un acuerdo no sólo dispondría una distribución imparcial o "justa" de los bienes disponibles, sino que permitiría un incremento neto de la cantidad de bienes. Esta simple intuición

auto-interesado de cada agente, pues aparece como un medio para la maximización. No sería racional conformarse con un estado sub-óptimo si otro óptimo es accesible[29]. Este razonamiento se hace evidente para el maximizador racional si, en vez de considerar únicamente las elecciones concretas a la vista de las estrategias que espera que los demás adopten en las situaciones del tipo del Dilema del Prisionero, considera también que su mismo modo de elegir afecta a las situaciones en las que puede esperar encontrarse. Porque el efecto de un modo "directamente maximizador" de elegir es desventajoso en este sentido. Se diría que el maximizador directo saca el máximo de sus oportunidades (desde su punto de vista individual), pero tiene muchas menos oportunidades que un agente dispuesto a cooperar. Si se comparan los resultados globales de las distintas *políticas* (la directamente auto-interesada y la cooperativa), resulta que la segunda es, en términos simplemente maximizadores, superior a la primera, porque "aunque el cooperador condicional se abstiene de sacar el mayor provecho a sus oportunidades, sin embargo se encuentra con oportunidades de las que el egoísta carece, así que puede esperar pagos superiores a los de éste"[30]. Las oportunidades a que se refiere Gauthier en este fragmento son, obviamente, las oportunidades de cooperar eficazmente con otro agente igualmente dispuesto.

Como conclusión de este (resumido) razonamiento, se establece la racionalidad de la cooperación. Es decir, cualquier agente perfectamente racional, ante la decisión sobre qué modo de interacción adoptar, *preferirá* el modo cooperativo por razones estrictamente auto-interesadas.

---

es un elemento básico de la ideología contractualista. Cfr. Gauthier, *MA*, pp. 114-115.

[29] Problema distinto —cuya solución es justamente el asunto que nos ocupa— es que la optimización exija un modo de interacción muy diferente al propugnado por el principio de la acción estratégica. Si en este caso la racionalidad prescribe maximizar la utilidad dadas las estrategias que espera que los demás elijan, en el caso de la cooperación (tras el acuerdo) la racionalidad demandará buscar —mediante la elección adecuada— un resultado óptimo dadas las estrategias acordadas de los demás. Como dice Gauthier (*MA*, p. 118), el principio de la racionalidad estratégica parece refutar al segundo principio, el de la cooperación. La tarea de la teoría del contrato moral es precisamente mostrar que esta impresión es incorrecta, que la cooperación es racional y que ambos principios pueden integrarse en uno solo.

[30] Gauthier, D., "The Incompleat Egoist", cit., p. 265.

La cooperación representa, por tanto, un modo de interacción que los agentes en el estado de naturaleza estarían dispuestos, en principio, a promover y a adoptar, sin necesidad de acudir a motivaciones externas a su auto-interés individual. Resta comprobar si la cooperación sería factible entre agentes maximizadores (no hay que olvidar que supondría una transformación en su concepción de la racionalidad) y, en caso de que lo fuera, cómo podría implementarse.

c) Excurso: estado de naturaleza y coordinación.-

Antes de iniciar el siguiente punto, donde discutiremos las posibilidades de la cooperación —esto es, el establecimiento de un contrato o pacto social mediante un proceso de negociación racional—, queremos aludir a una sugerencia de, entre otros, Hampton y Kavka, sobre la conceptualización del estado de naturaleza[31]. Se trata de la posibilidad de que la interacción en el estado de naturaleza no tenga exactamente, o no solamente, la forma de un Dilema del Prisionero, sino que se den mecanismos más o menos espontáneos de coordinación.

La coordinación surge cuando la estructura de la interacción permite que el resultado más beneficioso para todas las partes sea un punto de equilibrio, de forma que nadie incrementa su utilidad variando unilateralmente su estrategia. Existen muchas actividades coordinativas, por ejemplo la de dos remeros en un bote: es racional para cada uno remar si el otro rema, pero irracional si el otro no lo hace (pues el bote giraría en redondo); he ahí una actividad en la que dos resultados, ambos reman o ambos no reman, están en

[31] La idea de la coordinación en el estado de naturaleza es aludida en Hampton, J., *Hobbes and the Social Contract Tradition*, Cambridge, Cambridge U.P., 1988, p. 82 y ss. Gauthier dedicó un artículo ya clásico (fue pionero en la materia), titulado "Coordination", *Dialogue*, 14 (1975), pp. 195-224; reimpreso en *Moral Dealing*, cit., pp. 274-297.

equilibrio, y uno de ellos es preferido por ambos (ambos reman). En este caso, frente a lo que ocurría en el Dilema del Prisionero, la acción coordinativa es la estrategia racional y, una vez iniciada la actividad coordinada, la idea de defraudar carece de sentido.

Sin embargo, no todas las situaciones de coordinación se resuelven tan sencillamente como este ejemplo. En ocasiones no es tan fácil "coordinar estrategias", como puede verse en el juego de coordinación por antonomasia, que es el denominado "batalla de los sexos". Este juego plantea uno de los problemas que puede presentar la coordinación. Supongamos que en una pareja (Abel y Bárbara) uno de los dos (pongamos que el varón) es aficionado al fútbol, mientras que el otro es igualmente aficionado al teatro. A la hora de salir un domingo por la tarde, Abel preferirá ir al fútbol, y Bárbara preferirá ir al teatro; pero, hagan lo que hagan, ambos prefieren salir juntos. Asignando números del 0 al 5 a las utilidades derivadas de sus posibles acciones, podríamos estar ante una matriz de este tipo:

		B	
		va al fútbol	va al teatro
A	va al fútbol	5, 4	0, 0
	va al teatro	0, 0	4, 5

Ambos prefieren pasar la tarde juntos, pero Abel prefiere que ambos la pasen en el fútbol y Bárbara que ambos asistan al teatro. En este juego hay dos puntos de equilibrio entre los que no hay criterio racional alguno para decidir, pues ambos son óptimos de Pareto. En el caso de Abel y Bárbara, lo más probable es que dialogasen y (tarde o temprano) se pusiesen de acuerdo en ir, juntos, a uno u otro espectáculo. Como ya dijimos arriba, este acuerdo será —lejos de lo que ocurría en el Dilema del Prisionero— estable: ninguno de ellos tiene una razón auto-interesada para no ejecutar la acción requerida por

la estrategia conjunta.

En los casos en los que la comunicación es imposible, o en que por cualquier otro motivo no se puede seleccionar uno de los puntos de equilibrio (cuando hay más de uno), la coordinación dependerá de la intuición de los jugadores, o de ligerísimas diferencias en la intensidad de las preferencias[32]. Con todo, la coordinación en estos casos difíciles penderá en gran medida del azar, a pesar de ser querida (y tal vez buscada) por todas las partes.

La coordinación descubre una nueva dimensión en el análisis de la interacción pre-contractual. Porque si el estado de naturaleza presenta ámbitos donde la coordinación es posible, entonces cabría cuestionar que el resultado de la interacción natural fuese (o fuese necesariamente) la guerra de todos contra todos. Hampton, por ejemplo, defiende que si los agentes naturales hobbesianos no tuviesen el defecto de ser "cortos de miras", podrían no sólo coordinarse para realizar ciertas actividades mutuamente beneficiosas, sino incluso producir espontáneamente un marco para la cooperación.

La cooperación es inmediatamente costosa para cada agente (aunque colectivamente beneficiosa), como muestra el Dilema del Prisionero. No así la coordinación, que supone un beneficio directo para cada uno de los agentes participantes; por ello no es contradictoria con la racionalidad económica y puede aceptarse como una parte de la interacción natural que elevaría las expectativas de las personas antes del acuerdo.

Una interpretación maximalista de las posibilidades de la coordinación, combinada con un potencial surgimiento espontáneo de la cooperación[33], haría innecesario el acuerdo como mecanismo de entrada en la sociedad y en la

---

[32] Para explicar estas sutilezas, Gauthier introdujo, en 1975 (en "Coordination", cit., p. 289), el concepto de *salience* o preeminencia de un punto de equilibrio. Este concepto ha sido recientemente desarrollado por el Prof. Maarten Jansen (U. de Rotterdam), quien habla de *Focal Points* para explicar los casos en que los juegos de coordinación tienen resultados mejores que los que cabría esperar de la simple probabilidad de que los jugadores coincidieran. Según Jansen, en este tipo de situaciones se pone en marcha una faceta especial de racionalidad que conduce a seleccionar ciertos resultados que, debido a sus características —que escapan a una conceptualización uniforme—, "atraen" a la mayoría de los jugadores.

[33] Como sugiere Hampton en *Hobbes and the Social Contract Tradition*, cit., punto 3.1 y, con mayor refinamiento teórico R. Axelrod en *La evolución de la cooperación*, Madrid, Alianza, 1987.

moralidad. A nivel teórico, esta interpretación conecta con el denominado "programa de Hayek", que hemos mencionado en otro contexto[34].

Aunque no podemos profundizar aquí en esta dirección, resulta obvio que la interpretación "maximalista" aludida supone una alternativa al contractualismo como paradigma de justificación moral o política. Desde nuestro punto de vista, el contractualismo capta mejor las premisas básicas del liberalismo, y las desarrolla coherentemente, mientras que el "programa de Hayek" —de carácter más descriptivo que normativo— tal vez deja más lugar a ciertas ambigüedades. Por otro lado, aun concediendo que existiera un ámbito para la coordinación dentro de la interacción natural, éste podría ser limitado. Pese al beneficio, tanto colectivo como individual, que se derivaría de las actividades coordinativas, el resultado de la interacción libre seguiría quedando lejos de la optimalidad. Y mientras haya lugar para un incremento del beneficio, es decir, mientras no se haya alcanzado un óptimo, existe un fundamento racional para el acuerdo y la cooperación. Si esto es así, el contractualismo podría asimilar la hipótesis de la coordinación en el estado de naturaleza: aunque la coordinación permite aumentar la utilidad que reciben ciertos (o todos los) agentes respecto de un supuesto estado de no-interacción, sólo la cooperación permitiría alcanzar el potencial máximo de la interacción racional.

[34] Cfr. más arriba, cap. 3, nota 16.

3. Negociación racional y pacto

Lo que se dijo en el punto anterior sobre el estado de naturaleza puede resumirse en que la interacción natural tiene la estructura de un Dilema del Prisionero, cuya única solución —en caso de tener alguna— sería un acuerdo que permitiera la cooperación. Mediante la interacción cooperativa se garantizaría un resultado individualmente más beneficioso que el obtenido de la combinación de mutuos comportamientos estratégicos. Además, este resultado se aproximaría a un óptimo social. La racionalidad de la cooperación no puede ponerse en duda desde esta perspectiva.

Ahora bien, el hecho de que la optimalidad sea colectiva e individualmente deseable no es lo único que hay que tener en cuenta para establecer las bases de la cooperación. Entre los distintos óptimos, no todos son igualmente preferibles para todos los agentes. En el DP, por ejemplo, hay tres óptimos (los resultados 12,libre; libre,12 y 2,2) pero sólo 2,2 sería un óptimo factible tras una negociación entre ambos jugadores. Generalizando, el primer problema que la cooperación debe despejar es la elección de *un óptimo* de entre los muchos posibles.

Por otro lado, la cooperación es un modo de interacción que se caracteriza porque exige de cada agente la realización de una determinada estrategia pre-definida, perteneciente a la "estrategia conjunta"[35] (colectiva) más adecuada para producir el resultado óptimo. Por tanto, el acuerdo —desde el pacto entre los dos prisioneros, hasta el contrato social— consistirá en seleccionar el conjunto de estrategias, una para cada agente, cuya puesta en

[35] David Gauthier denomina *joint strategy* al conjunto de estrategias, una para cada agente, objeto de la negociación racional. Traducimos su término como "estrategia conjunta" (y con ello estamos, por una vez, de acuerdo con la traducción castellana de *MA*, de Alcira Bixio), aunque hubiera sido tal vez más explícito hablar de "estrategia común" o "estrategia colectiva". En este caso, hemos preferido la versión más neutra.

práctica simultánea conduce a que *todos* obtengan el resultado previamente determinado (por el mismo acuerdo).

a) Optimización, negociación y contrato social.-

Ahora podemos apreciar la importancia del concepto de optimalidad. El resultado de la interacción puede verse como un producto de estrategias, pero también como un conjunto de utilidades (una para cada agente). El Dilema del Prisionero nos ayudó a percibir esta doble perspectiva: en cuanto elector independiente, cada persona elige la estrategia directamente maximizadora; pero en cuanto miembro de un colectivo que ha decidido acordar una estrategia conjunta optimizadora, cada agente elegirá su estrategia teniendo en cuenta la utilidad. La esperanza común de alcanzar un resultado óptimo (que proporciona la mayor utilidad posible para cada agente) es el factor que "ata" a todos a la estrategia conjunta. Así, el propósito esencial del contrato es elegir un óptimo aceptable para todos. Parece lógico suponer que tal óptimo socialmente aceptable se identifica con la "preferencia social"; preferencia susceptible de ser representada en una adecuada función de utilidad social (que nos permitiría definir la optimización como un modo de "maximización de la utilidad colectiva", de un modo análogo a la maximización individual). Sin embargo, es sabido que el Teorema de Arrow prohíbe la definición de una función de utilidad social que representase adecuadamente las utilidades de los distintos individuos que integran el grupo. Según nuestra teoría, la elección de un óptimo social es, con todo, posible a través de una negociación racional, único mecanismo que tiene en cuenta los intereses individuales de *todas* las personas[36]. Sólo si se ha participado en una negociación perfectamente

[36] Gauthier justifica esta tesis en la p. 134 de *MA*. El argumento es que el utilitarismo pretende encontrar el resultado socialmente óptimo pasando por una ordenación social de preferencias. Es esa ordenación la que está prohibida por el Teorema de Arrow. La negociación permite "saltar" la función de utilidad social, y pasar directamente de las preferencias individuales a la elección de un resultado socialmente óptimo. Como dice Gauthier, esto pudiera parecer un "regate" al Teorema de Arrow, pero es un regate que funciona. De hecho, refleja bastante bien como se toman en la

racional puede cada agente tener la certeza de que la cooperación es beneficiosa para él. Y sólo la negociación asegura la selección de un óptimo social aceptable para cada individuo. Los demás mecanismos de "deducción" de la "función de utilidad (o bienestar, como a veces se dice) social" caen, efectivamente, bajo el Teorema de Imposibilidad de Arrow[37].

También podemos apreciar ahora con mayor claridad hasta qué punto la teoría del contrato ha de ser comprendida en su globalidad. La negociación es el núcleo de la teoría. Podemos imaginar que la negociación se desarrolla en una gran habitación con una mesa redonda en el centro, donde cada uno entra llevando en la cartera, y en la mente, el propósito auto-interesado de conseguir la mayor y mejor "tajada" posible (es decir, cada uno entra en ejercicio de su racionalidad maximizadora, y con la intención de seguir ejerciéndola durante el proceso negociador); pero también lleva cada uno el compromiso de llegar a un acuerdo con todos los demás, pues eso es precisamente lo que exige el auto-interés (la única razón para iniciar la negociación es la conciencia de que el resultado de la interacción natural es sub-óptimo; si es racional alcanzar un óptimo, entonces es racional no sólo negociar, sino hacerlo con la decisión de llegar a un acuerdo). Decíamos que la negociación es el núcleo de la teoría del contrato porque esa ficticia sala es el lugar donde se produce el paso del estado de naturaleza al estado social o, en los términos que venimos utilizando, el lugar donde el ejercicio de la interacción estratégica pone las bases para transformarse en interacción cooperativa. Como núcleo de la teoría, confluyen en la negociación los elementos cuya comprensión integral constituye la idea de un contrato moral. Porque en el proceso negociador es importante conocer cuál es el estatuto anterior de cada parte, es decir, qué ofrece cada uno de beneficioso al resto de los contratantes; cómo se desarrolla el proceso negociador mismo, y qué fórmula podría representar la actuación de los negociadores racionales; bajo qué condiciones es racional estar de acuerdo con

---

práctica las decisiones colectivas: mediante una negociación en que se discuten directamente los resultados en relación a las preferencias individuales. Cfr. también, sobre esto, Gauthier, "Bargaining and Justice" (en *Moral Dealing*, cit., pp. 187-206), pp. 190 y 199.

[37] Cfr. Gauthier, D., *MA*, cap. V, secc. 2.1 y 2.2, pp. 122-128.

el resultado de una negociación y actuar conforme al mismo; bajo qué condiciones será imparcial la acción cooperativa subsiguiente; etc. Todas estas cuestiones se entrecruzan: será racional aceptar el resultado de la negociación (y, por ende, cumplir el acuerdo) siempre y cuando el proceso negociador haya sido limpio o imparcial *y* esa imparcialidad se transmita a la cooperación; pero para que la cooperación aparezca como imparcial (o justa) no basta con que el proceso haya sido limpio, también habrá que tener en cuenta la situación inicial, porque nadie consideraría aceptable un acuerdo firmado bajo amenazas o coacción, así que será necesario establecer con claridad qué situaciones iniciales resultan inadmisibles desde un punto de vista racional (porque puedan asimilarse a las amenazas o coacciones). A su vez, poner límites a la interacción natural sólo tiene sentido en vista de la cooperación, pues ninguna persona aceptaría límite alguno a la libre interacción, salvo que fuera un medio necesario para obtener un beneficio mayor que el perjuicio causado por esa restricción.

En la negociación confluyen, en definitiva, todos los elementos de la teoría: problema de la situación inicial, principio de justicia distributiva, racionalidad de cumplir acuerdos, etc.

La estrategia de Gauthier, y la nuestra, consiste en analizar exclusivamente el proceso de negociación racional, dando temporalmente por sentado que es racional cumplir el acuerdo, y posponiendo la discusión sobre la situación inicial. Es decir, en lo que sigue, se aceptará sin discusión que las partes *poseen* algo como punto de partida de la negociación, y que esa posesión puede diferir de unos individuos a otros (este aspecto se restringirá en gran medida posteriormente); por otro lado, se aceptará también provisionalmente que los acuerdos se firman para ser cumplidos.

b) El objeto de la negociación.-

La cooperación es la respuesta racional al fracaso de la racionalidad económica. Cuando la persecución individual del auto-interés no es eficaz,

entonces hay que diseñar una estrategia conjunta que asigne a cada agente una estrategia optimizadora. Así se supera el fracaso de la razón como maximización directa. Mas esta superación implica un cambio de perspectiva sobre la racionalidad. Antes de considerar el objeto de la negociación propiamente dicho, hemos de apuntar en qué consiste ese cambio de perspectiva sobre la racionalidad, que es a la vez consecuencia y causa de la negociación. Las condiciones de la racionalidad estratégica enunciadas en el capítulo segundo (punto 2.d) han de ser reformulados. Se recordará que la primera condición decía que la elección de cada persona debe ser una respuesta racional a las elecciones que espera que otros hagan. Una respuesta racional significa una respuesta maximizadora de la utilidad, en un contexto de interacción no-cooperativa. Pero en un contexto cooperativo, lo que la condición exige no es tanto la maximización individual como la optimización[38]:

> "En ausencia de acuerdo sobre un resultado o conjunto de estrategias, es racional para cada persona tratar de maximizar su utilidad dadas las estrategias que espera que los otros elijan, mientras que en el contexto de un acuerdo es racional para cada uno intentar lograr un resultado óptimo dadas las estrategias acordadas de los demás."[39]

No podemos pretender, por el momento, otra cosa que enunciar esta modificación del requisito de la racionalidad estratégica. Sólo a la conclusión del argumento que ahora iniciamos aparecerá claramente la cooperación como una forma racional de interacción. La propia lógica de la negociación determinará las condiciones bajo las cuales un acuerdo (y, por ende, la cooperación) es racional, aceptable para las partes y, por tanto, susceptible de dar lugar a una modificación en la disposición de cada agente hacia la maximización directa.

---

[38] Cfr. Gauthier, D., *MA*, p. 117-118.

[39] Gauthier, D., *MA*, p. 118.

La optimización exigirá, esto sí podemos anticiparlo, un criterio o principio de elección que cumpla, para cada individuo, el papel que la maximización juega en la interacción no-cooperativa —es decir, el papel de orientación normativa de la decisión. Ese criterio o principio debe ser capaz de concretar la exigencia genérica de la cooperación, a saber, alcanzar un resultado óptimo. Porque, como es sabido, la cantidad de resultados óptimos que es posible alcanzar en una situación dada es prácticamente infinita. La optimalidad representa la "frontera de posibilidades de bienestar o utilidad" de un grupo, pero las distribuciones factibles de esa utilidad son indefinidas.

El objeto de la negociación es, por tanto, un principio que permita seleccionar un resultado óptimo determinado, y estará referido a la distribución de utilidades[40].

Las utilidades a las que se ha de referir el principio (aquellas cuya distribución ha de acordarse) tienen un límite objetivo, marcado por los beneficios de la cooperación (o "excedente cooperativo"). Con esto se indica que la utilidad de cada agente en el estado de naturaleza *no es negociable*, pues claramente sería irracional para cualquiera entrar en una empresa cooperativa que le va a suponer un coste neto. La dotación natural, cualquiera que sea, representa lo que cada agente "trae" a la mesa de negociación, no lo que en ella

---

[40] En términos de teoría de la negociación, tal como quedó fijada por Nash en 1950, el objeto de la negociación no es otro que buscar una "solución" a un problema de negociación, donde "una solución significa una determinación de la cantidad de satisfacción que cada individuo debe esperar obtener de la situación" (J.F. Nash, "The Bargaining Problem", *Econometrica*, nº 18, pp. 155-162; p. 155). Es interesante hacer hincapié en un aspecto presente en la definición de Nash que los teóricos contemporáneos no resaltan suficientemente (un olvido del que no se libra Gauthier, lo que es lamentable, porque se trata de un detalle pequeño, pero relevante para el contractualismo). Se habrá observado que el sentido de la solución de un juego de negociación está, para Nash, en *determinar* una cantidad (esperada) de satisfacción para cada individuo. Esta idea está incluso mejor expresada a continuación, cuando escribe que la solución debe ser "...una determinación de en cuánto debería valorar cada uno de esos individuos el hecho de tener esta oportunidad de negociar." Lo que queremos destacar aquí es que la solución de un problema de negociación tiene sentido porque, al permitir a cada individuo determinar exactamente el beneficio que obtendrá de la cooperación, posibilita la asignación de una concreta utilidad esperada a la situación de negociación misma (a la oportunidad de negociar). Esto es fundamental para una teoría que, como la teoría del contrato, se basa en la racionalidad individual: el hecho de que la oportunidad de negociar tenga una utilidad esperada para cada individuo (una utilidad que suponemos mucho mayor que la utilidad alternativa de "no aprovechar la ocasión") es el factor causal del contrato desde el punto de vista de la racionalidad individual. Y eso es todo lo que requiere el contractualismo para justificar la negociación y el pacto.

se discute.

El propósito del acuerdo es permitir la cooperación. Todas las partes conocen que la cooperación implica un beneficio, pues amplía el horizonte de sus utilidades posibles. No se discute, por tanto, ni si es racional cooperar (se parte de la base de que lo es de hecho), ni qué beneficio producirá la cooperación (producirá tanto como sea posible, lo que equivaldrá a la diferencia entre el *status quo* y el óptimo). Lo que se discute es cómo se ha de distribuir esa diferencia, es decir, el producto neto de la cooperación, dado que ésta sólo es posible contando con el asentimiento de todas las partes.

Cada parte —no olvidemos que se trata de negociadores perfectamente racionales y auto-interesados— intentará sacar la mayor ventaja del hecho de que su participación en la cooperación produce un beneficio para los demás. En una negociación entre sólo dos personas, la fuerza de esta arma es evidente. Si uno de los dos decide negarse a cooperar, el otro pierde cualquier posible beneficio. Si esto es así, cada negociador reclamará para sí *todo* el beneficio de la cooperación o, al menos, tanto como sea compatible con la obtención, por parte de los demás, de la utilidad que tenían en la posición inicial. Ahora bien, la simetría entre los negociadores implica que todos reclamarán lo mismo, y es a partir de ahí que hay que iniciar la negociación propiamente dicha. Al final de la cual tendremos una distribución del excedente cooperativo que habrá sido determinada por un principio racional de negociación.

c) Sobre el procedimiento de negociación.-

El procedimiento de la negociación determinará cómo se han de ir ajustando las demandas de las partes, inicialmente divergentes e incompatibles, hasta llegar a un acuerdo. Pero antes de explicar cuál es el procedimiento racional de negociación según Gauthier, nos gustaría hacer dos comentarios previos: sobre el modo de interacción de los negociadores, y sobre la necesidad misma de discutir el procedimiento de negociación, pese a que una teoría normativa como es la moral por acuerdo debería ser relativamente indiferente

a las características "técnicas" de su instrumental analítico[41].

Sobre el modo de interacción de los negociadores, hay que recordar una vez más —aun a riesgo de ser repetitivos— que la racionalidad que se pone en juego *en* la negociación es la simple racionalidad económica que nos sirve como paradigma elemental. La racionalidad "cooperativa" u "optimizadora" a que nos referíamos arriba habrá de regir la cooperación una vez establecida —pero eso ha de ser demostrado—, mas no interviene en el proceso negociador. La negociación representa el último caso de interacción competitiva; pero lo representa hasta sus últimas consecuencias. Cada negociador sólo aceptará, en la medida en que suponemos que es perfectamente racional, aquel acuerdo que maximiza su utilidad esperada. Cualquier otra solución de la negociación será considerada irracional y, por tanto, inaceptable.

El carácter estrictamente auto-interesado de la negociación garantiza que el principio de distribución elegido representa por igual los intereses de cada negociador. Ello se debe a que hay un reconocimiento mutuo de hallarse entre agentes perfectamente racionales, de modo que si para cada uno es racional maximizar su utilidad (es decir, maximizar la parte del excedente cooperativo que espera obtener), también lo es aceptar un grado de maximización semejante para los demás, pues sabe que si algunos agentes no alcanzan ese grado, impedirán el acuerdo (al igual que él mismo lo impediría si estuviera en su lugar). Esta simetría entre maximizadores racionales garantiza no sólo, como decíamos, que el principio distributivo representa los intereses de todos (es imparcial), sino además, que no puede ser recusado desde el punto de vista de la racionalidad instrumental, único criterio normativo que se ha introducido hasta el momento.

Este segundo comentario previo a la exposición del proceso negociador tiene por objeto satisfacer la posible duda sobre la pertinencia de entrar en los

[41] Este es el sentido en que se expresan, por ejemplo, Brian Barry, en *Theories of Justice*, Londres, Harvester-Wheatsheaf, 1989, p. 392; y Gauthier, D., "Uniting Separate Persons" (en Gauthier y Sugden [eds.], *Rationality, Justice and the Social Contract*, Ann Arbor, The University of Michigan Press, 1993, pp. 176-192), p. 178. Más abajo abundaremos en esta idea.

complejos vericuetos técnicos de la teoría de la negociación racional, cuando es evidente que este mecanismo analítico tiene, en la teoría del contrato social, un papel puramente heurístico. Cabe aducir que poco puede aportar a una empresa filosófica la detallada y fatigosa tarea matemática de postular axiomas, establecer teoremas, definir funciones y elaborar representaciones que, al fin y al cabo, iluminan muy tenuemente el problema en cuestión. El mismo Brian Barry, quien, en la primera parte de *Theories of Justice*, analiza en profundidad diversas propuestas de solución aportadas por matemáticos y filósofos para los juegos de negociación, reconoce que su refinamiento técnico no se corresponde con el filosófico. Pareciera que decir algo relevante sobre la justicia a base de intentar solucionar "racionalmente" un juego competitivo consistente en distribuir un bien escaso entre dos agentes con igual derecho al mismo, es un programa filosófico con poco futuro[42].

Sin embargo, Gauthier no renuncia al papel de la negociación en una teoría moral contractualista[43]. Ello significa tener que justificar la racionalidad de los principios de la justicia apelando a un proceso negociador que debe expresar la quintaesencia de la racionalidad individual. Y, lamentablemente, la respuesta de los matemáticos y economistas a esa apelación es difusa e incierta. La teoría de la negociación racional, que es relativamente joven[44], ha seguido rumbos variables y un tanto peregrinos a veces. Nash se esforzó en axiomatizar la teoría, de modo que los requisitos intuitivos básicos de una negociación racional quedasen claramente sentados, y el marco de discusión, establecido. Pero otros teóricos —cuando no el mismo Nash— pusieron el énfasis en aspectos difícilmente mensurables (y que se dan siempre en las negociaciones

[42] Cfr. Barry, B., *op. cit.*, p. 139.

[43] En este sentido, puede verse su "Rational Constraint: Some Last Word" (en Vallentyne, P. (ed.), *Contractarianism and Rational Choice*, Cambridge, Cambridge U.P., 1991, pp. 323-330), donde re-estructura por completo la teoría, asignando nuevos lugares a sus elementos, pero manteniendo un papel singular para cada uno de ellos y defendiendo, en especial, el papel de la negociación racional.

[44] La fecha (registrada y reconocida) de su nacimiento es 1950, cuando el matemático J.F. Nash publicó "The Bargaining Problem", en *Econometrica*, nº 18, pp. 155-162.

reales) tales como la capacidad de amenaza[45], la impaciencia[46], etc. Estos conceptos —así como las sucesivas modificaciones de los axiomas primigenios de Nash— han contribuido a lograr una precisión cada vez mayor en el análisis de las negociaciones reales, pero a costa de introducir variables que nada influirían en una negociación *ideal*. Por eso, Gauthier se vio obligado a desarrollar un modelo de negociación un tanto *sui generis* (lo que le ha valido, por cierto, un aluvión de críticas), tal que captase las peculiaridades requeridas por una teoría del contrato social.

La principal diferencia entre una negociación ideal y las negociaciones reales es que, en condiciones ideales, hay que suponer una estricta igualdad en el poder negociador de las partes. Ningún agente racional modificará sus demandas a causa de amenazas y esto, conocido por todos, elimina la posibilidad misma de que las amenazas se produzcan. Por lo que se refiere a la utilidad relativa del tiempo, también hay que descartarla, por definición, en una negociación ideal.

Estas suposiciones, que se pueden resumir en la simetría o igualdad en que se encuentran las partes, no implican la introducción de una "premisa moral"; son consecuencia de los postulados y premisas iniciales del argumento. Lo único ilegítimo, por injustificado, sería aceptar algún tipo de desigualdad en este punto.

Por otro lado, los procedimientos de negociación que los teóricos han discutido son muy diversos[47]. De hecho, el desarrollo de la teoría de la negociación racional ha consistido, básicamente, en introducir variantes y modificaciones de los procesos de decisión en la negociación. El filósofo moral se ve abocado a elegir uno de estos procedimientos; pero adoptar uno u otro de

---

[45] P. ej., Nash, J.F., "Two Person Cooperative Games", *Econometrica*, nº 21 (1953), pp. 128-140.

[46] P. ej., Rubinstein, A., "Perfect Equilibrium in a Bargaining Model", *Econometrica*, nº 50 (1982), pp. 97-109.

[47] Como ejemplo paradigmático del extraordinario grado de sofisticación alcanzado por algunos especialistas, puede contemplarse el complejísimo modelo de negociación de H. Moulin ("Implementing the Kalai-Smorodinski Bargaining Solution", *Journal of Economic Theory*, vol. 33, pp. 32-45.), sobre el que nos informa Ken Binmore en "Bargaining and Morality", cit., p. 154.

ellos puede determinar el resultado de la negociación. Esto no debería sorprendernos: al contrario de lo que sucede en ciertos juegos, cuyo resultado puede conocerse empíricamente, es decir, jugándolo (para contrastarlo después con lo que la teoría predijo), la negociación tiene un resultado incierto que impide ese tipo de "pruebas" empíricas. Cuál sea el resultado "justo" de una negociación racional es un problema puramente teórico. Es lógico, por tanto, que los componentes de cada modelo acaben influyendo en lo que se considera una "solución racional".

La elección entre los distintos modelos de negociación no es, para el filósofo moral, sólo una cuestión de adecuación técnica, como puede ser para el teórico de juegos. El filósofo moral debe considerar, además, la capacidad del modelo para tener en cuenta los factores relevantes en una "posición original" (y sólo esos factores) y para arrojar un resultado razonable a la vez que válidamente derivado de las premisas del argumento contractualista.

La necesidad de hacer esta elección, así como de profundizar y desarrollar el modelo de la negociación que hubiera de satisfacer las necesidades del contractualismo, condujeron a Gauthier a invertir no poco esfuerzo en su contribución a la teoría de la negociación racional. Esta contribución debe considerarse parte de una teoría contractualista: implícitamente, el contractualismo siempre supone una negociación entre agentes auto-interesados, cuyo resultado sería el acuerdo, el pacto social. Este mecanismo se hace explícito sólo entre los neo-contractualistas (y no entre los representantes clásicos de la tradición) debido a la generalización, en últimas décadas, del uso de la Teoría de Juegos. Gauthier decidió usar esta "herramienta" cuando aún no estaba definitivamente forjada, y se vio obligado a refinarla él mismo. Ahí podemos encontrar la causa histórico-genética de este comentario sobre el procedimiento de negociación. Si queremos añadir una causa lógico-material, sería que los modelos de negociación desarrollados por economistas necesitan cierta revisión para adecuarse a las exigencias impuestas por los estrechos márgenes de la racionalidad instrumental y la definición contractualista-liberal de los indivi-

duos[48].

El modelo de negociación de Gauthier está basado en el de Kalai-Smorodinski[49], aunque, como ponen de relieve Wulf Gaertner y Marlies Klemisch-Ahlert, "usa la idea de Zeuthen de una secuencia de concesiones sucesivas entre los jugadores"[50]. El modelo Kalai-Smorodinski resulta intuitivamente plausible y, dentro de lo que cabe, susceptible de una explicación "coloquial" (más abajo definiremos los conceptos fundamentales de la teoría de la negociación, lo que nos permitirá formulaciones más rigurosas). Se supone que el acuerdo permitirá a cada negociador una ganancia de utilidad. La diferencia entre la utilidad máxima que cada negociador puede recibir (si obtuviese todo el beneficio cooperativo) y la que ya posee en la situación inicial es su ganancia neta de utilidad máxima. Cada negociador racional intentará obtener su ganancia máxima, pero obviamente ese resultado (obtener todos su ganancia máxima) es inalcanzable pues cae fuera de la frontera de posibilidades de la sociedad. La solución de Kalai-Smorodinski consiste en definir el "grado de éxito" que cada negociador logra en su afán de alcanzar la ganancia máxima. Ese grado es simplemente la proporción de esa ganancia que finalmente consigue. Pues bien, el único punto en que esa proporción es igual para todas

---

[48] En contra de esta tesis, cfr. Ken Binmore, "Bargaining and Morality" (en Gauthier y Sugden [eds.], *Rationality, Justice and the Social Contract*, cit., pp. 131-156), p. 154. Binmore acepta el uso filosófico de una negociación hipotética, pero sostiene que la visión "ortodoxa" de la negociación —en vez de el modelo de Gauthier— podría cumplir este papel.

[49] Expuesto por primera vez en Kalai, E., y Smorodinski, M., "Other Solutions to Nash's Bargaining Problem", *Econometrica*, nº 43 (1975), pp. 513-518.

[50] Gaertner, W. y Klemisch-Ahlert, M., "Gauthier's Approach to Distributive Justice and Other Bargaining Solutions" (en Vallentyne, P. [ed.] *Contractarianism and Rational Choice*, cit., pp. 162-176), p. 171. Quizá no sea impertinente recordar que la idea de Zeuthen (un modelo negociador basado en concesiones mutuas) es anterior a la formulación axiomática de Nash, pues procede de su obra *Problems of Monopoly and Economic Warfare* (Londres, Routledge, 1930), donde propuso una solución para las negociaciones sobre los aumentos salariales. El modelo de Zeuthen se unió a la axiomatización de Nash y ambos fueron unificados por Harsanyi en "Approaches to the Bargaining Problem Before and After the Theory of Games" (*Econometrica*, nº 24, 1956, pp. 144-157) en lo que se viene considerando la "concepción ortodoxa" o "clásica" de la negociación. Gauthier se opone a esta concepción en el cap. V, punto 3.4 (pp. 146-150) de *MA*. El modelo de Zeuthen se asimila al de Nash-Harsanyi porque, pese a su afinidad con el de Gauthier por el uso de la idea de concesiones mutuas, su resultado resulta ser equivalente a la solución de Nash, basada en la maximización del producto de las utilidades.

las personas será la solución racional de la negociación. En una negociación racional, según Kalai-Smorodinski, todos los negociadores obtienen la misma proporción de la ganancia máxima que podrían obtener.

El primer artículo en que Gauthier explora las posibilidades de la negociación racional data de 1974[51], de forma que su punto de partida son las teorías de Zeuthen, Nash y Harsanyi. Pero ya las modifica en un sentido que le hará coincidir con la solución de Kalai-Smorodinski, publicada un año más tarde y que influye, como decíamos, en la formulación definitiva de *MA*. En "Rational Cooperation" Gauthier introduce la idea de "beneficio relativo" (que equivale en términos generales a la "proporción de la ganancia de utilidad" de Kalai-Smorodinski) y formula una condición para la cooperación racional[52] que se va a mantener hasta *MA*.

En "The social Contract: Individual Decision or Collective Bargain?"[53], Gauthier ofrece una versión completa de la teoría avanzada en "Rational Cooperation" y reconoce que su solución (pensada para la cooperación y la justicia) coincide con la de Kalai-Smorodinski:

> "Esta solución al problema del contrato social es una aplicación de un procedimiento general para la acción cooperativa que he desarrollado en otro lugar. Y este procedimiento es una generalización para *n*-personas de una solución formal al problema de la negociación desarrollada, independientemente de mí, por Kalai y Smorodinski."[54]

---

[51] "Rational Cooperation", *Noûs*, 8, (1974), pp. 53-65.

[52] Esta condición, que es el reverso de su conocido principio de concesión relativa *minimax*, se basa en que la "cooperación racional debe asegurar un resultado que haga que el beneficio relativo mínimo sea en mayor posible, o que ofrezca un beneficio relativo *maximin*" ("Rational Cooperation", cit., p. 56). Como quiera que el beneficio relativo *maximin* sólo se alcanza si se igualan los beneficios relativos de todos los negociadores, el principio de la cooperación se enuncia sí: "La cooperación es racional si y sólo si el resultado de la acción cooperativa proporciona un beneficio relativo igual máximo" (p. 57). Esta solución es idéntica a la de Kalai-Smorodinski.

[53] En Hooker, Leach y McClennen (eds.) *Foundations and Applications of Decision Theory, Vol. II*, Dordrecht, Reidel, 1978, pp. 47-67.

[54] Gauthier, D., "The Social Contract: Individual Decision or Collective Bargain?", cit., p. 58.

Con todo, la mayor originalidad de la solución de Gauthier al problema de la negociación habrá de esperar a *MA*, donde, combinando las ideas de beneficio relativo y concesión mutua, ofrece una contribución realmente novedosa en este campo. Analizaremos con detalle esta contribución en el epígrafe siguiente, mientras avanzamos en el presente algunas ideas sobre el *funcionamiento* del proceso negociador mismo.

Cada agente tiene, en la negociación, un punto de partida irrenunciable y unas expectativas (obtener *todo* el excedente cooperativo) inalcanzables. La negociación se inicia con una demanda por parte de cada jugador. El único requisito que ha de cumplir la demanda es no requerir que otros negociadores hayan de renunciar a parte de su utilidad inicial. Este requisito se cumple si todos demandan la totalidad del beneficio cooperativo. Esto es lo que denominaremos "demanda máxima" o "demanda racional". Lógicamente, es imposible satisfacer las demandas máximas de todos, de forma que cada uno debe hacer concesiones y renunciar a parte de lo inicialmente solicitado, hasta que las cantidades del beneficio cooperativo que los distintos individuos reciben se hacen compatibles[55].

Dos problemas surgen en este procedimiento. El primero es hasta dónde es racional ceder (nos ocuparemos de él inmediatamente). El segundo es algo más sutil, y se refiere a la demanda racional. En el párrafo anterior hemos hablado indiferenciadamente de la demanda máxima o racional de cada individuo. Sin embargo, cuál sea la demanda racional presenta ciertos problemas en el caso de negociaciones con más de dos personas implicadas.

En el caso de dos personas, la demanda racional se distingue claramente: cada uno solicitará *todo* el excedente cooperativo, es decir, todo el beneficio neto, producto de la cooperación. Pero este caso límite podría confundirnos: la demanda máxima no equivale siempre a *todo* el beneficio cooperativo; en palabras de Gauthier:

"Debemos tener cuidado para que la consideración de la negocia-

[55] Sobre el procedimiento de negociación, cfr. *MA*, p. 133.

> ción entre dos personas no nos lleve a malinterpretar la determinación de las demandas. En una situación que implique más de dos personas, no le estará permitido a cada uno reclamar todo el excedente cooperativo que podría recibir, sino sólo la parte del mismo a cuya producción hubiera contribuido. La demanda de cada persona está limitada por la amplitud de su participación en la interacción cooperativa. Porque si alguien demandase lo producido por la interacción cooperativa de otros, entonces esos otros preferirían excluirlo del acuerdo."[56]

Obsérvese que, teniendo en cuenta esta razonable limitación, la demanda racional será proporcionalmente más pequeña cuanto más grande sea el grupo de personas que han de cooperar. Sin embargo, este hecho no parece ser tenido en cuenta por Gauthier, quien, como veremos, concede una gran importancia a la demanda inicial en la determinación del resultado de la negociación[57].

---

[56] Gauthier, D., *MA*, p. 134.

[57] Robert C. Koons (Cfr. "Gauthier and the Rationality of Justice", *Philosophical Studies*, vol. 76, nº 1, Octubre, 1994, pp. 1-26; esp. p. 18-19) ha llevado hasta sus últimas consecuencias la tesis de Gauthier sobre el límite de la demanda racional en situaciones de más de dos personas, y su conclusión es que, conforme el número de personas crece, la regla de distribución (el principio de la justicia) y la precisa definición del estado de naturaleza son menos importantes. Es una conclusión devastadora para la teoría de Gauthier, una teoría cuya identidad reside justamente en defender un principio distributivo concreto y una concepción concreta de la situación inicial de negociación. Afortunadamente, el argumento de Koons descansa en la generalización de un principio económico (el principio de la utilidad marginal decreciente) de dudosa aplicación en este caso. Koons argumenta que conforme el grupo cooperativo crece, la proporción en que aumenta la utilidad de los miembros por la adición de un miembro más, disminuye. La "utilidad marginal" —valga el término— de cada nuevo miembro es menor, luego su demanda máxima se aproximará al punto de equilibrio de la negociación, con lo que el problema de distribuir a cada agente una parte de su demanda máxima será menos agudo. Este argumento —como la mayoría de los argumentos construidos desde premisas económicas— no considera que la "utilidad de la cooperación" —por mantener el lenguaje de la economía— es, comparada con los pagos del estado de naturaleza, tan grande, que esa distancia entre demanda máxima y equilibrio siempre será relevante. Si se nos permite una opinión muy personal, creemos que los argumentos económicos están diseñados para problemas de producción y distribución en los que siempre hay que tener presente una ajustada relación coste/beneficio; mientras que en el caso del contrato social, esa relación es, como argumenta Hardin, muy amplia (basta recordar que, en una interpretación puramente hobbesiana, el beneficio de la sociedad excede incluso el coste de soportar la coacción y el abandono de todos los derechos naturales). Las herramientas procedentes de la economía pueden ser útiles para la filosofía, pero quizá necesiten un nuevo calibrado antes de su uso.

d) El principio de concesión relativa *minimax*.-

El objeto de la negociación es alcanzar un acuerdo sobre la distribución del excedente cooperativo. El procedimiento, en dos fases, consiste en que cada parte avance una demanda racional y, posteriormente, inicie una serie de concesiones. La solución racional de la negociación dependerá de hasta qué punto estiman las partes que es racional ceder, pero ¿Cómo determinar la magnitud de esa concesión? Intentaremos contestar esta pregunta siguiendo la teoría de la negociación racional de David Gauthier. Para ello será necesario abandonar nuestro propósito de explicar el contractualismo moral con un lenguaje más o menos "común". La teoría de la negociación racional es un instrumento analítico que ha desarrollado una terminología propia (parte de cual ha ido apareciendo ya en lo anterior). Ahora precisaremos algunos conceptos, ofreciendo sus definiciones formales, para exponer con mayor rigor el modelo de Gauthier. Estas definiciones son comunes a los modelos de negociación más extendidos, pero intentaremos emplear el vocabulario concreto de Gauthier (bastante poco extendido, por cierto), para hacer más comprensible la explica-

---

Sobre el punto que discutimos, Russell Hardin mantiene una tesis completamente opuesta a la de Koons. En su artículo "Bargaining for Justice" (en E. F. Paul, *et al.* [eds.] *The New Social Contract*, Oxford, Blackwell, 1988, pp. 65-74) escribe que "el *status quo* anterior para cualquier individuo consiste esencialmente en no tener nada más que tiempo libre", y que "lo que tenemos que distribuir es virtualmente todo lo que tenemos en absoluto" (p. 73). Desde nuestro punto de vista, se trata de una visión bastante más acertada del contrato social que la de Koons. Sin embargo, no carece de problemas. Porque al suponer que la posición original es despreciable por referencia a los beneficios de la cooperación, Hardin niega que pueda tener cualquier influencia en la determinación de la distribución, por lo que —según él— la teoría de Gauthier daría como resultado un simple igualitarismo. Añade que "Gauthier supone que su teoría difiere del Principio de la Diferencia de Rawls en que tiene en cuenta el *status quo* anterior, mientras que Rawls pretende distribuir todo, incluido lo que era "mío" antes de la cooperación social. No parece que esto sea una diferencia significativa" (p. 73).

Casi es innecesario comentar que los análisis de Koons y de Hardin se basan en sendas exageraciones (por defecto y por exceso) sobre la relación entre la posición original y los beneficios de la cooperación. El análisis de Koons podría estar equivocado, por la razón que adujimos. En cuanto a Hardin, tal vez tiene razón en que la diferencia entre el principio de Rawls y el de Gauthier sería pequeña, pero eso no le autoriza a negar que sea significativa. Justamente se trata de una diferencia que, por mínima que sea, es significativa para cada individuo, porque representa para él la diferencia entre sentirse representado en el acuerdo originario o ser incapaz de ello.

ción subsiguiente.

La teoría de la negociación parte de la base de que las preferencias de cada negociador pueden representarse en sus funciones de utilidad[58], y que su comportamiento, supuesto que se trata de personas con una actitud no irracional hacia el riesgo, consiste en maximizar la utilidad esperada. Se supone, también, que la interacción no-cooperativa ofrece a cada negociador una utilidad determinada, que representa su posición inicial. La *posición inicial de negociación* será, pues, *un resultado*[59]: el resultado de la interacción no-cooperativa, que asigna una utilidad a cada jugador. Y esas utilidades constituyen la dotación inicial (lo que aportan a la mesa de negociación) de los futuros cooperadores. Este resultado, o posición inicial de negociación, puede representarse por un punto en un eje de coordenadas n-dimensional (donde n es el número de negociadores), que denominaremos "espacio de utilidades", pues se acepta convencionalmente que cada eje representa la cantidad de utilidad de cada uno de los agentes.

La *situación de negociación* se define como un conjunto no vacío de resultados posibles del proceso de negociación[60], representado en el espacio de utilidades por una figura cerrada, convexa, que denominamos *espacio de resultados*. La posición inicial de negociación es un punto dentro del espacio de resultados.

Un problema de negociación se soluciona si se encuentra un resultado

---

[58] Esta posibilidad fue axiomatizada por J. Von Neuman y O. Morgenstern, *Theory of Games and Economic Behavior*, Princeton, Princeton U.P., 1944 y 1947. La función de utilidad permite expresar a cada individuo sus preferencias en términos de utilidades numéricas (no meramente ordinales), de modo que tiene sentido hablar de diferencias de utilidad, o de "intervalos" de utilidad. Sin embargo, las comparaciones interpersonales de utilidad no son posibles, ya que la escala de utilidad para cada individuo es arbitraria (o, dicho de otra forma, la función de utilidad es independiente de las transformaciones lineales).

[59] No olvidemos que un resultado significa una utilidad para cada jugador. Por tanto, un resultado, en este caso, significa un conjunto de utilidades, una para cada individuo, derivada del juego de la interacción natural. En lo sucesivo, "resultado" se entiende siempre como un conjunto de utilidades.

[60] Se acepta que la situación de negociación incluye todos los resultados lógicamente posibles del proceso de negociación (dentro de los límites del espacio de resultados), contando entre ellos las loterías entre resultados. Esto permite que el conjunto de resultados sea compacto y convexo en sentido matemático; cfr. Nash, "The Bargaining Problem", cit., p. 158.

sobre el que todos los individuos estén de acuerdo. Por tanto, la solución de un problema de negociación consiste en elegir uno de los resultados comprendidos en la situación de negociación.

Como quiera que ya hemos anticipado las dos fases del procedimiento negociador, podemos pasar directamente a la cuestión sobre qué concesión es racional hacer. Pero antes de poder siquiera intentar una respuesta, es necesario especificar cómo se pueden medir las concesiones. La magnitud absoluta de una concesión, en términos de utilidad, es la diferencia entre la utilidad asociada al punto de demanda máximo y la utilidad del resultado propuesto como concesión. Pero esta magnitud no ofrece base alguna para relacionar las concesiones de los distintos negociadores, pues las utilidades no admiten comparación interpersonal[61]. Sin esta base, ningún negociador racional posee un criterio para valorar la "racionalidad" de su concesión.

Sin embargo, es posible introducir una medida de la *concesión relativa* que nos permita una comparación interpersonal. Para medir la concesión relativa tomamos el punto de la demanda máxima como la concesión 0 (pues si lo que un negociador logra su demanda máxima, no habrá cedido nada), y el retorno a la posición inicial de negociación se considerará una concesión completa. Así, se puede definir la magnitud relativa de *cualquier* concesión:

> "La magnitud relativa de cualquier concesión puede expresarse como la proporción en que está su magnitud absoluta con la magnitud absoluta de la concesión completa. Supongamos que la posición inicial de negociación proporciona a una persona una utilidad u^*, y su demanda racional le daría una utilidad $u^\#$, entonces, si accede a un resultado que le proporcione una utilidad u, la magnitud absoluta de su concesión es $(u^\# - u)$, de la concesión completa $(u^\# - u^*)$, y por tanto, la magnitud relativa de su concesión es $[(u^\# - u)/(u^\# - u^*)]$."[62]

---

[61] Cfr. Gauthier, D., *MA*, p. 134.

[62] Gauthier, D., *MA*, p. 136.

Es fácil ver que la concesión relativa, con un valor siempre menor que 1, establece una relación independiente de la escala de utilidad. Para todas las personas, el punto de demanda representa una concesión 0, y la posición inicial de negociación, una concesión 1. Así, sin introducir comparaciones interpersonales de utilidad, sí podemos comparar las concesiones relativas de las distintas personas en la negociación. Hecha esta comparación, cabe aplicar el principio de que, mientras no se llegue a un acuerdo, la persona que haya hecho una concesión relativa más pequeña, debe ceder[63]. Las concesiones serán progresivamente mayores, pero con el límite del acuerdo, esto es, en cuanto se alcance el acuerdo, ya no será racional ceder más. La generalización de esta idea da lugar al principio de la negociación según Gauthier:

> "Sostenemos que el principio establecería que, dado un conjunto de resultados, cada uno de los cuales requiere concesiones por parte de algunas o de todas las personas para ser seleccionado, entonces un resultado ha de ser seleccionado sólo si la concesión relativa mayor o *maximum* que requiere es lo más pequeña posible, o un *minimum*, esto es, no es mayor que la concesión relativa máxima requerida por otro resultado. Lo denominamos principio del minimum-maximum, o concesión relativa *minimax*."[64]

---

[63] Este principio fue introducido por Zeuthen (y luego un tanto olvidado por Von Neuman-Morgenstern y Nash). Se deriva inmediatamente de la igual racionalidad de las partes. Si el punto de acuerdo no ha sido alcanzado, ello significa necesariamente que, al menos uno de los negociadores, *debe* ceder. Es evidente que no sería razonable exigir una nueva cesión a quien ha hecho ya una concesión relativa *mayor* que los demás. Como se supone que las partes poseen la misma habilidad, conocimiento y capacidad negociadora, nadie puede esperar beneficiarse de la "estupidez" o de la "impaciencia" de otro; por tanto, quien haya cedido menos, reconocerá que —por decirlo así— "es su turno", y propondrá un resultado que le exija una concesión relativa mayor. Ahora bien, este aumento de la magnitud relativa de las concesiones de todos los negociadores ha de ser el menor posible (ahí se cifra la racionalidad intuitiva del principio de concesión relativa *minimax*), porque sería irracional hacer una concesión superior a la estrictamente necesaria para llegara a un acuerdo.

[64] Gauthier, D., *MA*, p. 137. Una explicación algo más clara (para principiantes), la encontramos en "Bargaining Our Way Into Morality: A Do-It-Yourself Primer" (en *Philosophical Exchange*, nº 2 [1979], pp. 14-27), p. 20: "Cada conjunto de acciones candidato para el acuerdo puede presentarse también como un conjunto de concesiones, una para cada persona. Cada uno de estos conjuntos debe tener un miembro mayor (la concesión *máxima* requerida para que se logre

La defensa de este principio se basa en que su base lógica está en el beneficio que cada parte espera obtener de la cooperación, es decir, apela al afán maximizador de las partes. Esta base lógica se cifra en las condiciones de la negociación racional, que reproducimos:

> "(i) *Demanda racional.* Cada persona debe reclamar el excedente cooperativo que le proporcione una utilidad mayor, con el límite de que nadie puede reclamar un excedente cooperativo si no ha participado en la interacción necesaria para producirlo.
> "(ii) *Punto de concesión.* Dadas las demandas según la condición (i), cada persona debe suponer que existe un punto de concesión factible que toda persona racional está dispuesta a concertar.
> "(iii) *Disposición a ceder.* Cada persona debe estar dispuesta a concertar una concesión en relación con un punto de concesión factible si su magnitud relativa no es mayor que la de la concesión más grande que él supone que cualquier persona racional aceptaría (en relación con un punto de concesión factible).
> "(iv) *Límites de la concesión.* Nadie está dispuesto a concertar una concesión en relación a un punto de concesión si ello no viene exigido por las condiciones (ii) y (iii)."[65]

---

el acuerdo sobre ese conjunto). Algún conjunto posible de concesiones debe tener un miembro mayor que *no* sea más grande que el miembro mayor de cualquier conjunto alternativo. Esta es la concesión *minimax* (la más pequeña, o mínima entre todas las posibles mayores, o máximas)."

[65] Gauthier, D., *MA*, p. 143. La explicación de estas condiciones se basa, como hemos dicho, en el beneficio que cada persona busca obtener de la cooperación. Gauthier explica que la condición (i) es una aplicación directa de la maximización de utilidad en el contexto de la negociación. La condición (ii) se sigue del hecho de que, para que exista la cooperación, debe haber un acuerdo. Negar que exista un punto de acuerdo racional es negar la posibilidad de la cooperación entre agentes racionales, pero es evidente para todos que la cooperación es beneficiosa y, por tanto, racional, luego deben estar dispuestos a llevarla a cabo. La condición (iii) expresa la igual racionalidad de las partes. Como maximizadores, todos tratan de minimizar su concesión, pero nadie puede esperar que otro vaya a ceder más si él mismo no está dispuesto a hacerlo. finalmente —prosigue Gauthier— la condición (iv) es de nuevo una aplicación directa del principio de la maximización de utilidad: dado que las condiciones anteriores permiten alcanzar un punto de acuerdo, ningún agente racional estaría dispuesto a hacer concesiones innecesarias.

Dado que el punto de concesión establecido por la condición (iii) representa la concesión relativa *minimax*, Gauthier sostiene que el conjunto de las cuatro condiciones han establecido el Principio de la negociación racional: "en cualquier interacción cooperativa, la estrategia conjunta racional está determinada por una negociación entre cooperadores en la cual cada uno avanza una demanda máxima y después ofrece una concesión no mayor, en su magnitud relativa, que la concesión *minimax*"[66].

Establecido el principio —solucionado, por tanto, el problema de la negociación racional y, consecuentemente, determinado el valor exacto de la cooperación para cada agente natural—, Gauthier retoma el argumento contractualista al explicar el papel que juega el principio de concesión relativa *minimax*.

Su papel, dice Gauthier, es triple. Primero, expresa el principio de la maximización de utilidad en el contexto de la negociación, es decir, representa exactamente lo que haría un maximizador racional, entre maximizadores racionales, en una negociación. Segundo, determina el contenido formal de una negociación racional: el acuerdo fijaría una estrategia conjunta que diese a cada persona una utilidad esperada exactamente igual a la asociada al punto de concesión. Por último, el principio de la concesión relativa *minimax* es el principio del comportamiento racional en la interacción cooperativa —la interacción basada en la estrategia conjunta acordada en la negociación. Porque en un contexto cooperativo, las acciones de cada persona no persiguen maximizar su utilidad, sino producir el resultado acordado, aquél que proporciona a cada persona una utilidad esperada no menor que la determinada por el principio de concesión relativa *minimax*.

Este tercer papel del principio es el más dubitable. Aún hay que demostrar que es racional actuar cooperativamente, conforme a lo que el acuerdo determina. Gauthier sostiene que, demostrado esto, se habrá acreditado el carácter moral del principio, porque, aplicado a la interacción cooperativa, constituiría una restricción imparcial en la persecución individual del auto-interés.

---

[66] *MA*, p. 145.

En relación con el triple papel del principio de concesión relativa *minimax*, es pertinente resaltar una vez más cómo se articulan —precisamente en este gozne del argumento— la maximización directa y la cooperación. Porque es el principio de actuación que seguirían maximizadores de utilidad *en* la negociación, pero a la vez se convierte en la guía del comportamiento cooperativo (no maximizador, sino optimizador). La teoría de la negociación racional expresa (quizá de una forma muy extensa y con un lenguaje poco familiar a la filosofía) una intuición difícil de explicar con palabras que no sean lugares comunes: la intuición de que el comportamiento moral, aparentemente opuesto al interés, representa el verdadero interés de cada persona.

e) La utilidad relativa *maximin*.-

El principio de concesión relativa *minimax* capta plausiblemente tanto la idea de un proceso negociador como la racionalidad maximizadora que las partes ejercitan en él. Esto lleva a Gauthier a seleccionarlo como principio de la negociación racional y, por ende, como regla de la cooperación. Sin embargo, es un principio que no capta suficientemente la idea de que la cooperación supone un beneficio para las partes. al poner el énfasis en la parte a la que hay que renunciar para alcanzar un acuerdo, puede dar la sensación de que la cooperación es costosa. Como sabemos, se trata justamente de lo contrario. Por ello, podemos leer el principio de un modo más natural, no fijándonos en la concesión relativa que cada persona hace en relación a su demanda racional, sino en el incremento relativo de utilidad que logra, por referencia al punto de no-acuerdo.

De hecho, este es un modo más "natural" de entender la negociación. Así la entendieron Kalai y Smorodinski y así comenzó su análisis el propio Gauthier.

Al intentar definir un principio de la negociación por referencia al beneficio —en vez de referido a las concesiones— llegaremos, de modo nada sorprendente, al mismo lugar por un camino distinto. Por eso, el principio de

la utilidad relativa *maximin* puede considerarse el reverso del principio de concesión relativa *minimax*. El último capta mejor la idea de una negociación racional, y el primero la de una cooperación mutuamente beneficiosa.

El análisis de la situación de negociación es el mismo. Cada negociador comienza con la utilidad que le ofrece la posición inicial, que llamamos u^*. La demanda racional será $u^\#$, que representa *todo* el excedente cooperativo. Aunque la magnitud absoluta del incremento de utilidad para cada punto no nos permite una comparación interpersonal, podemos definir la utilidad relativa que representa cada posible resultado. si la utilidad asignada al resultado es u, entonces la utilidad relativa de ese punto será $[(u-u^*)/(u^\#-u^*)]$. Se puede comprobar que la utilidad relativa es, para cada punto, la inversa de la concesión relativa (donde la concesión relativa es relativa es completa [1], la utilidad relativa es nula [0], etc.).

Para cualquier resultado, siempre habrá alguna persona que reciba el incremento relativo de utilidad menor; "yo entonces sostengo", escribe Gauthier, "que es una extensión natural de los requisitos de la maximización individual de utilidad que una negociación racional prescriba el resultado que maximice la utilidad relativa mínima"[67]. Esto se puede mostrar mediante un conjunto de condiciones similar a las condiciones (i) a (iv) que veíamos en el epígrafe anterior.

El principio de utilidad relativa *maximin* tiene el atractivo de coincidir (al menos formalmente) con la regla de decisión que presta su fundamento a los principios de la justicia de Rawls. Evidentemente, hablamos de "atractivo" en la medida en que esa regla de decisión y los principios derivados reflejen

---

[67] "Economic Rationality and Moral Constraints", *Midwest Studies in Philosophy*, III (1978), pp. 75-96; pp. 92-93. Una explicación más detallada del principio de la utilidad relativa *maximin* puede verse en "The Social Contract: Individual Decision or Collective Bargain?", cit., p. 56 y ss. Por otro lado, un artículo posterior, "Bargaining and Justice" (*Social Philosophy and Policy*, 2 (1985), pp. 29-47, habla de "beneficio proporcional" (*proportionate gain*), en vez de "utilidad relativa". Ese cambio terminológico puede calificarse como una rareza, una concesión a Zeuthen, Kalai y Smorodinski (con cuyas teorías debate en el artículo), o el inicio de una "vía muerta" sin continuidad tras "Bargaining and Justice".

nuestra natural "capacidad para un sentido de la justicia"[68]. La conexión del principio con la imparcialidad se acentúa si consideramos que, al igual que la concesión relativa *minimax* coincide, en muchos casos con la menor concesión relativa igual; así la utilidad relativa *maximin* tiende a coincidir con la mayor posible utilidad relativa igual. Se trata, por tanto, de un principio de distribución estrictamente proporcional, lo que coincide con la concepción pre-teórica más común de la justicia.

Estos atractivos intuitivos de la versión "positiva" del principio de la negociación no deben hacernos olvidar que se trata del reverso del principio de concesión relativa *minimax*. Son principios estrictamente equivalentes; dos perspectivas sobre la misma noción teórica. Por lo tanto, lo dicho sobre la segunda puede trasladarse también a la primera.

f) Devenir y crítica del principio.-

Ya indicamos, al hablar del procedimiento de negociación, que la incorporación de estas ideas al proyecto de filosofía moral liberal de Gauthier se produjo hacia 1974, fecha de su "Rational Cooperation". Al comienzo, la idea de una negociación como mecanismo para la elección de los principios de la justicia debió nacer como alternativa a las teorías de la justicia de Harsanyi y Rawls, según las cuales esa elección puede asimilarse a una decisión individual bajo ciertas estipulaciones. Este sentido se observa ya en "Rational Cooperation" (p. 59) y, transformado en un argumento completo, en varios artículos posteriores[69]. Lo que en "Rational Cooperation" era una intuición relativamente poco precisa, deudora sobre todo de la Teoría de Juegos

[68] Cfr. Rawls,J., *Political Liberalism*, Nueva York, Columbia U.P., 1993, p. 19.

[69] Especialmente en "Justice and Natural Endowment: Toward a Critique of Rawls's Ideological Framework", *Social Thory and Practice*, nº 3, 1974, pp. 3-26; y "The Social Contract: Individual Decision or Collective Bargain?", cit.

—"puesta de moda" por R.D. Luce y H. Raiffa, autores de *Games and Decisions* (Nueva York, Wiley 1957)— se convirtió pronto en una original contribución a la joven Teoría de la Negociación Racional.

Gauthier desarrolló inicialmente el concepto de utilidad relativa y el principio "positivo" de la negociación (completamente formulado en 1978, pero basado en el trabajo de 1974). Mas la noción de concesión relativa está tan estrechamente conectada con la de beneficio relativo que es fácil suponer que se incorporó rápidamente; especialmente teniendo en cuenta que capta mejor la naturaleza del proceso negociador. Esta idea aparece publicada por primera vez en 1979[70]. En aquellos primeros trabajos, Gauthier ya intentaba relacionar la negociación y la justicia. Pero esto no suponía ninguna novedad: desde sus orígenes, el propósito de la teoría de la negociación racional se cifra en la determinación de una distribución equitativa o justa de algún bien. Quizá por ello, la versión de Gauthier no despertó especial interés, salvo en ámbitos muy especializados.

Pero en *MA* la teoría de la negociación racional se integra en una teoría moral contractualista, y el principio de la negociación es adoptado como principio moral. A partir de este momento, las críticas al modelo de negociación de Gauthier se multiplican, las deficiencias y debilidades del principio son detectadas, y se cuestiona su adecuación como principio moral. Esa es la pequeña historia que queremos narrar (al menos en sus episodios más interesantes) ahora[71].

---

[70] Cfr. "Bargaining Our Way into Morality: A Do-It-Yourself Primer", cit.

[71] Nos vamos a centrar en las críticas más interesantes, que conciernen al principio de la negociación. Otros críticos, como Peter Danielson ("The Visible Hand of Morality", *Canadian Journal of Philosophy*, vol. 18, nº 2, junio 1988, pp. 357-384; esp. p. 363 y ss.), Robert Sugden ("Rationality and Impartiality: Is the Contractarian Enterprise Possible?", en Gauthier y Sugden, *Rationality, Justice and the Social Contract*, cit., pp. 157-175) o el español J.C. Bayón Mohino (*La normatividad del derecho: deber jurídico y razones para la acción*, Madrid, Centro de Estudios Constitucionales, 1992, p. 166 y ss.), han intentado recusar la teoría desde un punto de vista global (dando la sensación de que aceptan la solución técnica de Gauthier), arguyendo, por ejemplo, que el principio de concesión relativa *minimax* no logra individualizar un sólo resultado racional, o que desde el punto de vista de la racionalidad individual serían aceptables resultados "aproximados" al determinado por el principio, con lo que su función como principio de justicia quedaría en cuestión, etc. Dejaremos de lado este tipo de críticas, primero, porque son demasiado vagas y escasamente concluyentes y, segundo, porque se pueden considerar más bien críticas a la teoría

Desde el punto de vista de los teóricos de juegos, la teoría de la negociación de Gauthier es "heterodoxa". El mejor argumento en su favor es que, para casos límite, su resultado coincide con el resultado de la aplicación del principio de negociación de Nash[72]. Consecuentemente, la crítica más inmediata se cifra en mostrar que, para otros casos, el resultado difiere del

---

moral gauthieriana como conjunto, no tanto al modelo de negociación.

[72] Gauthier compara su modelo con la teoría de la negociación de Zeuthen-Nash-Harsanyi (que toma como versión "ortodoxa") en la sección 3.4 del cap. V de *MA* (p. 146 y ss.). Allí muestra un ejemplo en que el resultado arrojado por el principio de concesión relativa *minimax* difiere de la solución de Nash y —sólo por casualidad o por la deliberada elección del ejemplo— resulta intuitivamente más plausible. La diferencia entre el resultado de Gauthier y el de Nash se puede explicar con una analogía (la cual nos evitará entrar en aspectos técnicos). Nash postula que la solución de la negociación es un resultado tal que el producto de las utilidades asignadas a cada persona sea máximo. En cierto sentido, este resultado se asemeja a la elección de un principio de justicia utilitarista, esto es, trata indiferenciadamente las utilidades de todas las personas, pues la cantidad a maximizar está en relación con la "utilidad total" distribuida. Si hay un resultado que "distribuye" más utilidad, el procedimiento de Nash lo elegirá, aunque la distribución sea poco equitativa (pues la prioridad de su principio es la maximización de lo que podemos llamar "utilidad conjunta"), mientras que el procedimiento de Gauthier seleccionará un resultado más acorde con nuestras pre-concepciones sobre la equidad, aunque para ello deba renunciar a optimizar la cantidad total de utilidad distribuida. Gauthier justifica su opción arguyendo que en una negociación, cada parte no está interesada en maximizar ninguna medida conjunta, sino en maximizar *su* propia utilidad resultante (cfr. especialmente Gauthier, D., "Bargaining and Justice", en E.F. Paul *et al.* [eds.], *Ethics and Economics*, Oxford, Clarendon, 1985, pp. 29-47; también es interesante el comentario de Wulf Gaertner y Marlies Klemisch-Ahlert, "Gauthier's Approach to Distributive Justice and Other Bargaining Solutions", cit., p. 166).

La opción de Gauthier tal vez es razonable, pero representa un punto débil de su modelo, hábilmente detectado y analizado por, entre otros, Russell Hardin en "Bargaining for Justice", cit., pp. 68-69, y Martín Diego Farrell en *La Filosofía del Liberalismo*, Madrid, Centro de Estudios Constitucionales, 1992, pp. 80-86. Según estos autores, es ilógico renunciar a parte del beneficio global que podría ser logrado a cambio de realizar una distribución como la que produciría una negociación racional. En ese tipo de situaciones (sub-óptimas) la solución se impone por sí misma: calcular la distribución según el principio de concesión relativa *minimax*, luego producir toda la utilidad posible y, tras efectuar la distribución determinada por el principio, dividir el resto a partes iguales o proporcionales. Este sería un método para no "desperdiciar" utilidad, que es —siempre según esta crítica— la consecuencia de la aplicación del principio de Gauthier.

La defensa de Gauthier en este caso apelaría al individualismo que su modelo postula. Su interés es encontrar un principio de justicia que no pueda ser recusado por ningún individuo *ex post*, y no tanto determinar un resultado económicamente óptimo desde un punto de vista neutro. Para ello, la estructura de la cooperación debe ser tal que cualquiera reconozca que no podría haber "jugado mejor sus cartas" en la negociación. Dada la información disponible y las condiciones de la posición original, Gauthier defiende que "la mejor jugada" consiste en aceptar como resultado el fijado por el principio de concesión relativa *minimax*. En cierto modo, la "jugada" que proponen críticos como Hardin cuenta implícitamente con un conocimiento mayor del que la teoría concede como plausible a las partes en la posición original.

"ortodoxo".

En concreto Wulf Gaertner y Marlies Klemisch-Ahlert han mostrado que la solución del modelo de Gauthier no está bien definida para los casos que implican a más de dos personas[73]. En estos casos, el mantenimiento de la solución requerida por el principio de concesión relativa *minimax*, exige la formulación de modificaciones y nuevos axiomas al conjunto de condiciones de la negociación definidas por Nash. Gaertner y Klemisch-Ahlert ofrecen una axiomatización del modelo de Gauthier, de tal modo que funcionase para los casos de *n*-personas[74]. Su análisis concluye que, para un sub-conjunto considerable de negociaciones de *n*-personas, las condiciones establecidas por el modelo de Gauthier se pueden mantener sólo si la solución que se adopta está regida por una variante del principio de la utilidad relativa *maximin* (cfr. epígrafe anterior), que denominan "solución lexicográfica *maximin* sobre incrementos relativos de utilidad"[75]. Esta solución parece ser la única que permite individualizar un resultado óptimo que distribuya utilidades de tal modo que la ganancia menor de utilidad que alguien ha de aceptar, sea la mayor posible. El resultado así seleccionado mantiene el fundamento racional de la concesión relativa *minimax*; de hecho, Klemisch-Ahlert lo considera una mejora y adaptación del modelo de Gauthier.

El principio lexicográfico del incremento de utilidad relativo *maximin* formulado por Klemisch-Ahlert como principio generalizado de la negociación fue aceptado por Gauthier en "Rational Constraint: Some Last Word"[76], ante las deficiencias de la formulación de *MA*. Sin embargo, en el mismo artículo reconoce que, tanto el principio de concesión relativa *minimax* como el principio lexicográfico de Klemisch-Ahlert, pueden presentar otras inadecuaciones más profundas. Algunas de ellas son señaladas por diversos críticos y

---

[73] Cfr. "Gauthier's Approach to Distributive Justice and Other Bargaining Solutions", cit., p. 172.

[74] Cfr. Gaertner, W. y Klemisch-Ahlert, M., art. cit., p. 173-174.

[75] Gaertner, W. y Klemisch-Ahlert, M., art. cit., p. 174.

[76] En Vallentyne, P. (ed.), *Contractarianism and Rational Choice*, Cambridge, Cambridge U.P., 1991, pp. 323-330; p. 325.

contrarrestadas por Gauthier en "Moral Artifice". Pero algunas otras le han llevado, más recientemente, a cuestionar la validez de su propio modelo.

Jean Hampton es la autora que en más aprietos puso, en un primer momento, al principio de concesión relativa *minimax*[77]. Sin el despliegue teórico-matemático de otros críticos, Hampton usó diversos ejemplos para mostrar claramente que el principio defendido por Gauthier producía resultados contrarios a nuestras más elementales convicciones sobre la justicia, tales como la idea de proporcionalidad en las retribuciones. Lo más chocante es que el principio *minimax* tiene la apariencia de ser más proporcional que igualitario, de modo que la evidencia de lo contrario suponía un reto al espíritu del principio. No obstante, al no apoyar su crítica más que en ejemplos, Hampton reconoce no tener una posición segura desde la que recusar la tesis de Gauthier. En algún momento incluso propone varios principios de justicia, en pie de igualdad, y confiesa que todos ellos "suenan" equitativos.

En nuestra opinión, la crítica de Hampton es suficientemente replicada por Gauthier[78], al hacer notar que un principio de proporcionalidad sólo puede ser deducido y considerado plausible, como regla de distribución, porque se olvida que el principio seleccionado lo es *de una negociación*. Hampton no tiene en cuenta, por ejemplo, que el hecho mismo de concertar un acuerdo (el consentimiento) se ha de considerar una contribución al mismo, de la que no debe extrañar que se deriven retribuciones que, lógicamente, no se corresponden proporcionalmente con la aportación económica de las partes. El principio *minimax* sirve —escribe Gauthier[79]— para definir la conmensurabilidad de la aportación implicada por el consentimiento de las partes. El resumen de la

---

[77] Cfr. "Equalizing Concessions in the Pursuit of Justice: A Discussion of Gauthier's Bargaining Solution", en Vallentyne, P. (ed.), *Contractarianism and Rational Choice*, cit., pp. 149-161.

[78] Cfr. "Moral Artifice", *Canadian Journal of Philosophy*, vol. 18, nº 2, junio 1988, pp. 385-418; pp. 390-391.

[79] Cfr. "Moral Artifice", cit., p. 394.

replica de Gauthier es, por tanto, que Hampton se refiere a un principio de distribución ya establecido, que nada tiene que ver con un acuerdo fundante:

> "Es verdad que, en muchas situaciones, la distribución de beneficios es independiente de un acuerdo. Pero tales situaciones no son cooperativas en el sentido en que *MA* se refiere a la cooperación. Más bien son situaciones típicas de mercado. El tipo de interés de mercado para una inversión hace que la cooperación, y por tanto un principio de distribución como la concesión relativa *minimax,* sean innecesarios."[80]

Pese a la convincente réplica que Gauthier opone a la sugerencia de Hampton, "Moral Artifice" deja traslucir la conciencia de que el principio *minimax* no posee demasiada solidez. En páginas posteriores, Gauthier ha de lidiar los problemas de la determinación del resultado de la negociación en interacciones multipersonales —por entonces aún no era conocido el principio lexicográfico de Klemisch-Ahlert— y concluye reconociendo que el principio *minimax* quizá necesite ser refinado, con la condición de preservar su fundamento en la igual racionalidad (maximizadora) de las personas.

Aparte de las disputas más o menos técnicas sobre el modelo de negociación y su principio, hay dos ideas en "Moral Artifice" que nos parecen destacables, ambas dirigidas contra sendas malinterpretaciones del capítulo V de *MA*. La primera es la insistencia de Gauthier en que el principio *minimax* no ha de ponerse en relación con nuestras intuiciones morales o con cierto "sentido de la justicia", sino exclusivamente con un análisis de la negociación racional. La segunda es que la teoría de la negociación ha de ser entendida en un marco teórico mayor, que es un argumento contractualista cuyas premisas (básicamente el individualismo y la racionalidad instrumental) deben mantenerse presentes. El primer comentario puede leerse como una defensa del principio *minimax* frente a sus posibles competidores; el segundo como una defensa del

[80] *Ibidem.*

papel genérico de la negociación para el caso de que el principio sea, con todo, insostenible.

Se puede afirmar que, ya en 1988, tan solo dos años tras la publicación de *MA*, la crítica al principio de la negociación había extendido dudas sobre el mismo —dudas reflejadas en las palabras de Gauthier. Pues bien, cinco años más tarde aquellas dudas dejaron paso a nuevas certezas, expresadas, por ejemplo, en el siguiente fragmento:

> "¿Debo abandonar la concesión relativa *minimax*? Mi opinión en el presente es esta. El argumento del capítulo V de *MA* no puede mantenerse en su forma actual. Como mucho, puede tener un valor heurístico al presentar la idea de la concesión relativa *minimax* como nexo entre la racionalidad y la moralidad. Pero el verdadero trabajo de defender la concesión relativa *minimax* como un resultado de la negociación, si —como aún creo— fuera defendible, requiere un argumento diferente. Requiere argumentar que en las circunstancias del contrato social, la concesión relativa *minimax* coincide con la solución de Nash."[81]

Esta rotunda afirmación es la respuesta de Gauthier al análisis de los juegos de negociación no-cooperativos presentado por Ken Binmore, y que concluye con la tesis siguiente:

> "Estoy seguro de que Gauthier hace bien en convertir la negociación hipotética en una noción fundamental para una teoría del contrato social. También estoy seguro de que es correcto no confundir el problema introduciendo consideraciones éticas en el análisis del procedimiento de la negociación. [...] Pero no veo ninguna buena razón para no usar la teoría ortodoxa de la

[81] Gauthier, D., "Uniting Separate Persons", en Gauthier y Sugden (eds.), *Retionality, Justice and the Social Contract*, cit., pp. 176-192; p. 178.

> negociación para predecir el resultado del proceso de negociación. Como he intentado defender, la teoría ortodoxa está plausiblemente basada en modelos que tratan de minimizar los supuestos ficticios necesarios, mientras que nada parecido puede decirse de la solución Kalai-Smorodinski, ni del análogo multi-personal de Gauthier."[82]

El argumento de Binmore resulta convincente por diversas razones —la principal de ellas es que, basándose en el modelo de Rubinstein[83], logra efectivamente minimizar los supuestos contra-fácticos de la negociación racional. Ante él, Gauthier flexibiliza su tesis aceptando el resultado del modelo ortodoxo. Es cierto que en "Uniting Separate Persons" intenta una tímida defensa de la tesis de que, en las circunstancias del contrato social, el resultado de Nash coincide con el principio *minimax*[84]; pero la conclusión general de estas discusiones está mejor captada, creo, por Brian Barry cuando afirma que "dado el uso que queremos hacer una teoría de la negociación en este libro (y

---

[82] Binmore, K., "Bargaining and Morality", cit., p. 154.

[83] El modelo de Rubinstein elimina la necesidad, que existía en el de Nash, de tener en cuenta la "habilidad" negociadora de las partes o, para no complicar el modelo, *suponer* que todas las partes tenían una habilidad semejante. Rubinstein introduce, a cambio, el tiempo. El paso del tiempo en el proceso negociador logra explicar muchas limitaciones al proceso que antes sólo podían justificarse mediante reglas artificialmente impuestas. Considerar el paso del tiempo en la formalización matemática de un proceso negociador crea muchas dificultades, por supuesto; pero estas se evaporan cuando, en el caso límite, el tiempo tiende a cero; y así "el caso límite del juego de Rubinstein *minimiza* las ficciones necesarias para explicar el comportamiento de quienes participan en el juego de negociación. En particular, las reglas del juego dejan a los jugadores sin ningún motivo obvio para violarlas. Por tanto, no es necesario fingir hipotéticamente castigos exógenamente determinados para el caso de infracción de las reglas. Ni tampoco es necesaria la hipótesis de un mecanismo exógeno para mantener los compromisos. Se trata, por consiguiente, de una notable vindicación de la intuición de Nash de que el único (*unique*) resultado en equilibrio, en el caso límite en que el tiempo tiende a cero, es una *solución de la negociación de Nash* ponderada." (Binmore, K., art. cit., p. 150).

[84] Gauthier afirma que "el principio de concesión relativa *minimax* exige que el contrato social distribuya el excedente cooperativo óptimamente y con iguales expectativas de beneficio relativo, siempre que esto sea compatible. La solución de Nash exige algo semejante dado que el contrato social resulta de una negociación que es simétrica en los resultados racionales individuales. Suponiendo esta simetría, puedo defender la exigencia de la concesión relativa *minimax* sin separarme del camino de la ortodoxia representado por Binmore y Rubinstein" (art. cit., p. 179).

es el mismo que Gauthier quiere hacer), no necesitamos realmente una teoría del proceso de negociación diseñada para explicar cómo calculan las personas sus ofertas iniciales y cómo llegan desde ellas a un acuerdo. No nos interesa el proceso a través del cual se alcanza el resultado. De hecho, lo que necesitamos es exactamente lo que ofrece la solución de Nash. Es decir, necesitamos un resultado previsible de la negociación que refleje el poder negociador de las partes"[85]. Al hacer estas afirmaciones Barry demuestra una clara comprensión del papel de la negociación en la teoría del contrato social. Pese al lenguaje utilizado, que proviene de la Teoría de la Elección Pública, lo que está en juego no es tanto la distribución de bienes públicos, cuanto la instauración de un criterio racional de justificación de instituciones, reglas y comportamientos sociales. La negociación y el acuerdo subsiguiente tienen un papel *orientador*, y ni la desviación, frecuente en la práctica, del resultado predicho por la teoría, ni la aceptación más o menos razonable de convenciones o tradiciones injustificables desde un punto de vista racional e imparcial, cuestionan ese papel. Gauthier ha insistido en ello —repitiendo ideas anteriores— en su última réplica a los críticos[86], escribiendo:

> "Conforme las personas llegan a ser conscientes de la idea de un contrato social, vienen a considerar justificables las convenciones, instituciones y prácticas sociales sólo en la medida en que pueden representarse como resultados plausibles de tal contrato, concebido como una negociación racional *ex ante* sobre los términos de la interacción social. Tal vez se sometan racionalmente a convenciones que consideran injustificables, dado que existe una expectativa de adherencia general a las mismas y, en algunos casos, una imposición. Pero con el tiempo las expectativas

[85] Barry B., *Theories of Justice*, cit., p. 392.

[86] En concreto, Gauthier contesta a una cuestión de Sugden referente a la racionalidad de desafiar las convenciones establecidas y, por tanto, a la racionalidad de poner en práctica las demandas de la negociación racional, que Gauthier identifica con la justicia (cfr. "Rationality and Impartiality: Is the Contractarian Enterprise Possible?", en Gauthier y Sugden [eds.], *Rationality, Justice and the Social Contract*, cit., pp. 157-175; p. 170).

cambian, y la obligatoriedad de convenciones injustificables se reconoce cada vez más como arbitraria. Así, sólo aquellas convenciones e instituciones que puedan representarse como soluciones plausibles de un contrato social, resultarán sostenibles, porque sólo ellas mantendrán (si ya existían) o ganarán (si no existían inicialmente) el apoyo voluntario de los miembros de la sociedad. Lo que es racional aceptar en la interacción del mundo real se irá conformando a lo que sería racional acordar al determinar *ex ante* los términos de la interacción. La brecha entre lo que es racional y lo que es justo se cerrará"[87]

Desde el punto de vista que representa esta tesis, la discusión acerca de la

[87] Gauthier, D., "Uniting Separate Persons", cit., p. 180. La última frase de esta cita quizá parezca algo injustificada, pues —dada nuestra interpretación del contractualismo moral— no hemos puesto demasiado énfasis en el hecho de que el resultado de una negociación racional puede considerarse justo si se cumplen ciertas condiciones tales como la igual racionalidad de las partes, un procedimiento imparcial y una posición inicial de negociación no-coactiva. Gauthier defiende la tesis de la justicia del resultado de la negociación en "Bargaining and Justice" (*Social Philosophy and Policy*, 2 (1985), pp. 29-47; reimpreso en Gauthier, D. *Moral Dealing*, cit., pp. 187-206). Este artículo, concebido en parte como una refutación del modelo de teoría de la justicia como decisión individual (Rawls, Harsanyi), explicita la razón por la que una negociación racional representa mejor los principios de la justicia y se adecua mejor al papel evaluador y orientador (de la práctica social) propio de tales principios. Los siguientes fragmentos transmiten lo esencial de la posición de Gauthier sobre este tema: "Suponemos que los principios de la justicia constituyen la solución de un problema de negociación apropiadamente especificado, y así [...] suponemos que los principios de la justicia pueden ser representados como la maximización del beneficio proporcional mínimo esperado por las partes. Los principios de la justicia son principios cuyo fin es maximizar el beneficio proporcional mínimo. [...] Si, de hecho, G es la solución racional al problema de la negociación, entonces un resultado que maximice el beneficio proporcional mínimo debe ser racionalmente aceptable desde el punto de vista de todos los individuos. Cada persona, reconociendo la igual racionalidad de todos, considera racional adecuar su propuesta y concesión de modo que se maximice ese mínimo. Así, un acuerdo basado en la solución G es completamente imparcial, no tanto porque se abstraiga de los intereses de los individuos implicados, sino porque reconoce cada uno de sus intereses de un modo racionalmente aceptable desde el punto de vista de cada persona. En la negociación, logramos la imparcialidad entre personas reales al tomar en serio la distinción entre ellas." (p. 202); "El objetivo de la moralidad no es maximizar cierta cantidad análoga al bien individual; por el contrario, el objeto de la moralidad es el modo en que se distribuyen entre los individuos los beneficios que la sociedad hace posibles. La moralidad relaciona esta distribución con un acuerdo entre esos individuos. Y así, la racionalidad de la decisión moral está asegurada, no asimilándola a la racionalidad de una elección individual en condiciones de riesgo, sino construyéndola a imitación de la racionalidad de una negociación." (p. 205).

precisión técnica de los principios, se relativiza considerablemente. La conclusión que cabe extraer es que una teoría de la negociación racional que capte las ideas y premisas centrales de la interacción natural tal como quedó definida y consiga producir un principio racional para la cooperación, juega un papel relevante en una teoría contractualista. Que el modelo de la negociación sea más o menos "ortodoxo"; más o menos fiel a los axiomas de Nash; más o menos basado en ejemplos de negociaciones reales, etc., es un problema que tal vez interese a economistas y matemáticos, pero ocupa un segundo plano para el filósofo. Desde el punto de vista del contractualismo moral, nuestra única preocupación es velar por que el procedimiento de negociación no incluya subrepticios pre-supuestos morales (como sucede en el caso del principio proporcional defendido por Hampton), ni postulados incompatibles con las premisas que especificamos en el capítulo segundo. En la medida en que el modelo "ortodoxo" respeta estas restricciones, es admisible como modelo ideal de negociación. Esta es, básicamente, la tesis de Gauthier en "Uniting Separate Persons". Esta tesis es fortalecida por la evidencia de que, en las circunstancias del contrato social, el resultado de varios modelos de negociación coincide, a grandes rasgos, con el resultado del principio de concesión relativa *minimax*. Con ello, la pretendida imparcialidad del principio (que Gauthier necesita para mostrar, después, que incorpora un sentido de la justicia) parece que puede ser defendida. En definitiva, el modelo de Gauthier —pese a sus carencias— refleja perfectamente el resultado producido por la interacción racional en un contexto de negociación ideal y, en este sentido, mantiene su validez como fuente de un principio de cooperación racional capaz de demandar el cumplimiento universal *ex post* entre personas racionales.

Podemos aceptar, por tanto, que el resultado de la interacción natural no es tanto una "guerra de todos contra todos", sino un acuerdo sobre un principio de cooperación racional que fija los términos de la interacción social[88]. Se

[88] En lo sucesivo, aceptaremos generalmente que el contenido de este principio es la concesión relativa *minimax*, aunque sepamos que este contenido está mejor representado —al menos según los últimos textos de Gauthier— por el principio lexicográfico de la utilidad *maximin* de Klemisch-Ahlert y, de modo más general, por el resultado de la negociación según el modelo-límite de

trata de un resultado racional en la medida en que es fruto de las decisiones libres de individuos maximizadores de utilidad. Queda así demostrado que es racional concertar un acuerdo que abra paso a la cooperación. Ahora bien, una vez suscrito el pacto, ¿es racional cooperar? Piénsese que hasta el momento hemos considerado únicamente interacciones entre maximizadores directos de la utilidad. El mismo criterio racional que lleva a los jugadores del Dilema del Prisionero a confesar, o a los contribuyentes a defraudar, es empleado en la negociación para "sacar la mayor tajada". Hasta aquí —y pese a las frecuentes confusiones entre interacción cooperativa y negociación— sólo hay un tipo de comportamiento: maximizador. Pero, una vez concluida la negociación y concertado el pacto, ha de entrar en juego un nuevo tipo de comportamiento, definido por la cooperación y la optimización. ¿Es racional y, sobre todo, es posible, ese cambio de comportamiento? Esta es la cuestión que trataremos de responder —siguiendo, como siempre, la propuesta de Gauthier— en el siguiente punto.

Rubinstein. Seguiremos utilizando el principio (y el término) *minimax* por una razón de comodidad, amparados en que es *equivalente* a los resultados "ortodoxos".

4. La racionalidad como maximización restringida

El problema que discutiremos en este punto —la racionalidad de cumplir los acuerdos pactados— es el suscitado por el incorregible (y prudente) discurso del Tonto (*Foole*) en el capítulo XV del *Leviatán*. Recordemos que el *Tonto* no cuestiona que sea racional a veces concertar acuerdos, ni que se llame justicia al cumplimiento de los mismos e injusticia a su quebrantamiento. Lo que cuestiona es que sea racional, siempre, ser justo.

Si aceptamos que la negociación y el pacto, al producir un esquema cooperativo, exigen de cada agente la adherencia a los principios o reglas (mutuamente beneficiosas) limitadoras de su auto-interés, hemos de admitir también una doble dimensión en nuestra discusión presente: la racionalidad de cumplir con los pactos en general, *y* la racionalidad de ser moral (esto es, de adherirse a reglas intersubjetivas que frecuentemente demandan del agente actuar en contra del criterio de la maximización). A esta doble finalidad encamina Gauthier su "teoría del cumplimiento"; en palabras de Kraus y Coleman:

> "Una teoría del cumplimiento debería explicar por qué los agentes racionales cumplirían la estrategia conjunta que negociaron. Al pactar una estrategia conjunta, los agentes han acordado someterse a principios normativos, quizá morales. Así, el problema de explicar por qué los agentes racionales se plegarían a una estrategia conjunta se convierte en el problema de determinar el componente motivacional de una teoría moral —esto es, la parte de una teoría moral que explica por qué es racional actuar conforme a principios morales."[89]

[89] J.S. Kraus y J.L. Coleman, "Morality and the Theory of Rational Choice", *Ethics*, 97 (julio 1987), 715-749, p. 719.

Pero una teoría del cumplimiento es un artefacto bien complejo de diseñar. Hobbes no halló, entre los materiales proporcionados por la racionalidad individual, elementos suficientes para construirla. La posición del *Tonto* sólo pudo ser atacada con un argumento que recurría a la amenaza y la coacción. En vez de con una teoría del asentimiento racional, Hobbes replica al *Tonto* con una teoría de la obligación política.

Gauthier confía en los medios de la Teoría de la Decisión Racional y los Juegos para poder vencer al *Tonto* en su propio terreno. Promete ofrecer una defensa de la racionalidad de cumplir acuerdos sobre la misma base que esgrime el egoísta o el escéptico moral para defender su incumplimiento: la maximización individual de utilidad.

David Gauthier espera encontrar una solución al dilema de la cooperación dentro del marco teórico del análisis económico de la interacción. Y es una esperanza compartida por muchos de quienes han reflexionado sobre estos dilemas, si bien suficientemente débil como para dudar del éxito de la empresa. John Watkins puede ejemplificar esta convicción común:

> "Es natural esperar que la reflexión sobre las funestas consecuencias de la persecución racional del auto-interés en las situaciones tipo Dilema del Prisionero, debería motivarnos a cambiar el egoísmo por el moralismo o *cuasi*-moralismo; es seguro que un egoísta racional se daría cuenta de que todos, él incluido, saldrían mejor parados si el resultado [cooperativo] se considerara superior al [equilibrio]. Sí, pero en ausencia de un *Volkgeist* o entendimiento común que se "moralizase", arrastrando tras de sí a los individuos egoístas, *todos* se convertirán en moralistas sólo si cada uno de los individuos egoístas decide cambiar (y ningún individuo moralista decide volverse egoísta). Y cuando consideramos qué incentivo tendría un individuo egoísta para cambiar, los

resultados anteriores sugieren una respuesta bastante desalentadora."[90]

Los "resultados anteriores" a los que se refiere Watkins son los que cabe esperar de un análisis del Dilema del Prisionero basado en los criterios "ortodoxos" de la teoría bayesiana de la decisión. Desde ese punto de vista, cualquier intento de "moverse hacia el moralismo" (o la cooperación) choca invariablemente con la densa resistencia del criterio de maximización. Parece que ninguna consideración de auto-interés que pudieramos calificar como "interna" a la situación misma de los prisioneros es capaz de superar el dilema. Sin embargo, satisfacer la objeción del *Tonto* requiere superarlo; requiere hacer ver que es racional, no sólo negociar y pactar, sino cooperar de hecho sobre la base del acuerdo concertado. Y esto hay que hacerlo con la fuerza suficiente para *motivar* al egoísta o al escéptico moral. A este efecto semi-imposible, Gauthier despliega un argumento sólo relativamente nuevo[91]. Como primera aproximación al mismo podemos decir que sostiene que, si bien es individualmente racional tratar de maximizar el auto-interés en cada situación estratégica —y, por tanto, concertar o no concertar, cumplir o no cumplir, los pactos según aconseje el cálculo prudencial—, cabe también plantear una decisión, no sobre la estrategia a emplear en una situación determinada, sino sobre la "política" a seguir en ese tipo de situaciones. La tesis es que, para una agente racional, debe ser posible reflexionar (y tomar decisiones) sobre el *modo* mismo en que las decisiones son tomadas, es decir, sobre la disposición racional que

---

[90] Watkins, J., "Second Thoughts on Self-Interest and Morality", en Campbell, R. y Sowden, L. (eds.), *Paradoxes of Rationality and Cooperation. Prisoner's Dilemma and Newcomb's Problem*, Toronto, The University of British Columbia Press, 1985, pp. 59-74; p. 73.

[91] Quizá la fuente más inmediata de la propuesta de Gauthier sea la visión de Kurt Baier sobre la motivación moral (Cfr. p. ej., *The Moral Point of View*, Ithaca, Cornell U.P., 1958, p. 310); pero también los teóricos de juegos han sugerido ideas como la de "meta-juego" para intentar solventar el Dilema del Prisionero mediante la ficción de permitir que las partes elijan, no la estrategia a seguir, sino un "plan de acción" o un "modelo de racionalidad" al que ha de ajustarse la estrategia subsiguiente. No obstante, muchos elementos del argumento de Gauthier sí son originales; como lo es, sobre todo, el sentido radical de su defensa de la posibilidad de "elegir disposiciones" (que se opone a la opinión mayoritaria de los especialistas en Teoría de la Decisión) y la incorporación de la maximización restringida como elemento esencial de una teoría contractualista.

se va a adoptar. Esto no supone el abandono de la concepción instrumental maximizadora de la racionalidad. Todo lo contrario, los agentes son capaces de evaluar las distintas opciones (modos de decisión) que se abren ante ellos precisamente porque tienen un criterio de evaluación, que es la maximización de la utilidad esperada. Pues bien, Gauthier sostiene que, siguiendo el objetivo normativo de maximizar la utilidad esperada, un agente racional optaría por una disposición a la cooperación; disposición que determinaría las elecciones sucesivas. Desde el punto de vista de la nueva disposición, que Gauthier denomina "maximización restringida" (en lo sucesivo MR), resulta *racional* cumplir lo pactado —sea cual sea la matriz de pagos de la situación estratégica concreta— si se dan ciertas circunstancias (concernientes tanto al acuerdo mismo como a la actitud previsible del agente con quien se interactúa). Lo que se ha producido es, por tanto, un cambio en el paradigma de la racionalidad: el "maximizador restringido" encuentra racional cumplir (condicionalmente) los pactos suscritos *independientemente* de los pagos esperados, a diferencia del maximizador directo (en lo sucesivo MD), quien sólo cumplirá lo pactado cuando el cumplimiento sea, además, la estrategia que le proporcione mayor utilidad esperada.

Este "cambio de paradigma" será explicado como una extensión natural de la racionalidad maximizadora y permitirá dar razón del cumplimiento del pacto, al mostrar que es racional cooperar. En este sentido, la idea de la MR juega un papel importante en la teoría de la negociación —pese a que su deducción es lógicamente independiente de ella—, pues sólo un pacto que va ha ser cumplido llegará a establecerse. De alguna forma, la creencia en la racionalidad del cumplimiento (al menos como posibilidad) está presente en el inicio de la negociación, como condición necesaria (al menos subjetivamente) para la misma. La racionalidad de la negociación y el pacto no dependen solamente de la imparcialidad del procedimiento, sino también de que sea factible el cumplimiento. Por tanto, la racionalidad de la negociación y del pacto va ha quedar supeditada al resultado de su defensa.

El ejemplo del Dilema del Prisionero puede ilustrar, una vez más, lo que decimos. Se recordará que nada cambiaba en el juego por el hecho de que ambos presos acordaran guardar silencio: la matriz de pagos era idéntica, las

motivaciones de las partes también; el resultado, previsible. Tal situación conduce a eliminar la posibilidad de pacto entre agentes racionales, conscientes de que no hay diferencia entre pactar o no[92]. Sólo si es posible prever el cumplimiento (al menos con cierto grado mínimo de probabilidad) será racional iniciar una proceso negociador y pactar los términos de la cooperación.

Pero, por otro lado, nadie estaría dispuesto a cumplir un acuerdo que le otorgara menos de lo que cualquier agente racional puede exigir en una negociación imparcial (esto es, un incremento de utilidad *maximin*). Por lo tanto, la racionalidad de cumplir dependerá, a su vez, de que la estrategia conjunta que ha de ponerse en práctica sea la requerida por el principio *minimax*.

El cumplimiento —y, por ende, la maximización restringida— representa, entonces, el paso de un acuerdo hipotético a la restricción real (pues recordemos que en la negociación los agentes son libres, en cuanto maximizadores; y no restringen, sino que tratan de avanzar, sus intereses particulares). La negociación, último momento de la interacción estratégica, da paso a la interacción cooperativa. Obviamente, esta "conversión" de los maximizadores

[92] Este análisis se ha aplicado, por ejemplo, a la situación de la guerra fría (que proporciona un ejemplo especialmente claro). Imaginamos dos potencias nucleares, *A* y *B*, ambas gastando gran parte de su presupuesto nacional en proseguir una carrera de armamentos rigurosamente inútil (pues ya ambas poseen armamento suficiente para destruirse mutuamente varias veces). Es evidente que ambas mejorarían sus balanzas de pagos (y su bienestar) si ambas eliminaran todo el armamento nuclear (supongamos que mantendrían un equilibrio estratégico empleando solamente armas convencionales). Existen, por tanto, las condiciones para un acuerdo. ¿Por qué no se produce? Simplemente porque nada cambiaría. Imaginemos la situación (paralela al razonamiento de los presos). Ambas potencias acuerdan eliminar su arsenal nuclear. El estado mayor de la potencia *A* razona "si *B* elimina su armamento, es nuestra oportunidad de, manteniendo el nuestro, lograr nuestro objetivo de invadirlos y vencerles para siempre; si no lo hace, habremos hecho mejor en conservar nuestro arsenal, para seguir en una posición equilibrada; así pues, haga lo que haga *B*, lo mejor que podemos hacer es incumplir el pacto y no eliminar nuestro arsenal". Pero el estado mayor de *B* haría el mismo razonamiento. Y, lo que es más importante, ambas potencias conocen con antelación este "teatro de operaciones" post-contractual, de modo que, con muy buen criterio, ni siquiera intentan un pacto que, con toda seguridad, sería inútil. Ahora bien, una vez concluida la guerra fría —y somos conscientes de que aquí intentamos una extensión del ejemplo que comporta numerosas y bastante problemáticas derivaciones— es posible pensar que el acuerdo daría paso a un nuevo paradigma normativo que ya no tendría en cuenta las posibles ventajas estratégicas del incumplimiento, sino que restringiría el abanico de acciones racionales a aquellas de acuerdo con lo pactado.

en cooperadores supone la negación del mismo concepto de racionalidad con el que se inicia el argumento contractualista, y ha suscitado un aluvión de críticas de parte de los partidarios de la concepción bayesiana de la racionalidad. Iremos aludiendo a las diferentes críticas al paso de nuestra exposición del argumento de Gauthier, pero queremos anticipar una somera clasificación de las mismas (se nos disculpará que el contenido de algunas de ellas no quede suficientemente claro, debido a que aluden a puntos del argumento sobre la MR que no hemos mencionado aún).

El primer grupo de críticas —el más numeroso y mejor articulado— proviene de los que podemos llamar "bayesianos ortodoxos", que cuestionan la posibilidad de "salir" del paradigma de la racionalidad como maximización de la utilidad esperada. Desde su punto de vista, el cumplimiento del pacto revela una preferencia individual por la cooperación, de modo que no se puede hablar, en rigor de restricción del interés, sino de un cambio en los intereses del agente (y, por tanto, en su función de utilidad). Estos críticos se niegan a admitir que sea posible compatibilizar una estructura de interacción tipo Dilema del Prisionero con una respuesta cooperativa en el marco de análisis de la Teoría de la Decisión Racional.

Un segundo grupo de críticos se centran en los inverosímiles supuestos que Gauthier ha de introducir para defender la racionalidad de adoptar la disposición MR. Gauthier ha de suponer que los seres humanos somos "translúcidos", en el sentido de que es posible, para otro agente, reconocer nuestras disposiciones con cierto grado de claridad situado entre la transparencia total y la opacidad completa. Se trata, indudablemente de una hipótesis *ad hoc* recusada por diversos comentaristas.

Hay una tercera línea de crítica que acepta los términos generales del argumento de Gauthier, pero niega que su conclusión se siga del mismo. A cambio, proponen algún modelo alternativo de conformidad, menos restrictivo que el defendido por Gauthier. En esta línea se sitúan Bayón Mohíno, Kraus, Coleman y P. Danielson, entre otros.

Por último, existe un tipo de crítica radical, que cuestiona la posibilidad del tipo de transformación subjetiva que la adopción de la MR requiere. Según estos críticos, es la misma idea de "adoptar una disposición" la que está

desencaminada, pues no está en manos del agente el cambiar discreccionalmente su "disposición" racional. Este tipo de crítica se suele mezclar y confundir con las formuladas por los que hemos denominado "bayesianos ortodoxos". Pero hay diferencias. E. McClennen puede tomarse como ejemplo: niega la posibilidad de "adoptar una disposición", pero a cambio ofrece otro mecanismo de restricción del auto-interés. Coincide con los bayesianos en la crítica, pero se aparta de ellos porque cree que es posible —si bien no por el camino que intenta Gauthier— justificar el cumplimiento de los acuerdos pactados.

Tal vez todas estas posibles líneas de crítica, pueden resumirse en el dilema que con tanta claridad ha formulado Jung Soon Park[93]: o bien es irracional cumplir el acuerdo y sólo cabe imponerlo coactivamente (Hobbes), o bien hay que suponer que los individuos son ya sujetos morales dispuestos de antemano a cumplir el acuerdo que conduce al mantenimiento de una sociedad justa (Rawls).

Este dilema manifiesta en toda su crudeza la dificultad del proyecto de Gauthier en este punto; porque justamente su intento se cifrará en demostrar que cabe superar el dilema partiendo de su "cuerno" hobbesiano. Veamos cómo es ello posible.

a) ¿Qué exige la cooperación?.-

A la conclusión del punto anterior aceptábamos que un agente racional estará dispuesto a negociar hasta asegurar la selección de una estrategia conjunta optimizadora e imparcial. "Una estrategia optimizadora imparcial será aquella de la que cabe esperar que —dadas las estrategias previsibles de los demás— produzca un resultado (aproximadamente) imparcial [*fair*] y óptimo: un resultado cuyos pagos de utilidad se acerquen a los del resultado cooperati-

[93] Cfr. Park, J.S., *Contractarian Liberal Ethics and the Theory of Rational Choice*, Nueva York, Peter Lang, 1992, p. 159.

vo, tal como quedan determinados por el principio de concesión relativa *minimax*"[94]. Pues bien, podemos definir a un agente cooperativo[95] como el que adopta una estrategia optimizadora imparcial.

La optimización rige la acción de la persona dispuesta a cooperar, lo cual supone una nueva lectura de la primera condición de la racionalidad estratégica.

Se recordará que en el capítulo II (punto 2.d) enunciábamos las condiciones de la racionalidad, la primera de la cuales (condición A) reza que "la elección de cada persona debe ser una respuesta racional a las elecciones que espera que otros hagan". Un agente cooperativo, acepta la siguiente lectura de esa condición (las condiciones B y C no varían):

> A': La elección de cada persona debe ser una respuesta imparcial optimizadora a las elecciones que espera que otros hagan, siempre que esa respuesta sea posible; en otro caso, su elección debe ser una respuesta maximizadora de la utilidad.[96]

Por lo tanto, lo que exige la cooperación es cumplir la condición A' en vez de la primitiva condición A de la racionalidad estratégica —es decir, llevar a cabo y aceptar una transformación de las condiciones de la racionalidad. Lo primero que debemos preguntarnos es qué implica exactamente el cumplimiento de esta condición de la racionalidad modificada.

Una primera respuesta nos lleva a diferenciar claramente el comportamiento basado en la condición A' (MR) tanto de la maximización a largo plazo

---

[94] Gauthier, D., *MA*, p. 157. Sobre el uso del adverbio "aproximadamente" entre paréntesis, Gauthier explica que "en muchas situaciones una persona no esperará que los demás hagan exactamente lo que el principio de concesión relativa *minimax* exige, de modo que será imposible elegir una estrategia cuyo resultado fuese completamente imparcial o totalmente óptimo. Pero suponemos que [una persona justa] seguirá estando dispuesta a interactuar cooperativamente, en vez de no-cooperativamente."

[95] Gauthier comienza a utilizar, desde el comienzo del capítulo VI de *MA* el término "persona justa" [*just person*] para referirse a los agentes dispuestos a cooperar. El significado de su término no difiere del más neutro "agente cooperativo".

[96] Gauthier, D., *MA*, p. 157.

como de cierto tipo de transacción mutuamente beneficiosa.

Gauthier distingue con especial énfasis la MR de "una buena política para alcanzar la maximización a largo plazo"[97]. Es una distinción que, en efecto, necesita ser enfatizada, ya que el modo más natural e inmediato de entender la MR es como una "estrategia global, o a largo plazo", consistente en enmascarar el objetivo de la maximización directa a fin de obtener oportunidades de las que un "maximizador sincero" se vería excluido. Un argumento de este tipo ("es racional comportarse como una persona justa ya que, a largo plazo, resulta más rentable") tal vez sería convincente para el escéptico moral (que podría transformarse en un "egoísta ilustrado", como sugiere Baier), pero no implicaría restricción alguna[98]; el escéptico moral no tendría ningún motivo para no sacar ventaja de cualquier posible "excepción" circunstancial a esa política de calculada hipocresía. Sin embargo, la adopción de la disposición MR impediría ese tipo de actuación. Como dice Gauthier, la MR no es la MD con su más efectiva máscara[99]. Muy al contrario, la MR se inscribe en un esfuerzo —sostenido por los teóricos de juegos desde el planteamiento mismo de las paradojas tipo Dilema del Prisionero o problema de Newcomb— tendente a "elaborar un nuevo concepto de racionalidad que no sacrifique el principio de maximización de la utilidad, y al mismo tiempo estimule el comportamiento cooperativo"[100].

Otra forma equivocada de aproximarse al concepto de MR sería entender que la restricción en la persecución del auto-interés es una forma de transacción mutuamente beneficiosa. Cada uno espera un beneficio como consecuencia de la restricción impuesta a los demás y una pérdida como consecuencia de la restricción que él mismo soporta; pero estima que aquel beneficio supera esta pérdida.

---

[97] En este sentido, puede verse *MA*, p. 170. en este punto, Gauthier es secundado, entre otros, por E.F. McClennen; cfr. "Constrained Maximization and Resolute Choice", en E.F. Paul *et al.* (eds.), *The New Social Contract*, cit., pp. 95-118; p. 102.

[98] Cfr. Park, J.S., *op. cit.*, p. 154.

[99] *MA*, p. 169.

[100] J. Barragán, "Las reglas de la cooperación", *Doxa*, 6 (1989), pp. 329-384; p. 347.

Esta explicación tiene cierta apariencia de plausibilidad, en la medida en que, efectivamente, cada persona aumenta su utilidad respecto a la situación de no-cooperación gracias a que *todos* se ajustan a una estrategia conjunta. Sin embargo, no es un enfoque correcto porque la adopción de una disposición no puede depender de un beneficio cuya causa es externa al agente que delibera sobre la adopción de la disposición en cuestión. La idea de una transacción mutuamente beneficiosa explicaría el cumplimiento relacionándolo con una modificación en la matriz de pagos post-contractual: si tras el acuerdo *todos* adoptan la MR, entonces, el resultado de actuar uno mismo cooperativamente produce una utilidad esperada superior al resultado de seguir una estrategia individual maximizadora. Pero el *modo* de interacción no ha cambiado en absoluto. Se ha producido, simplemente, un cambio en la situación. Por tanto, la idea de la transacción no justifica adecuadamente la adopción de la MR[101].

Los dos malentendidos que acabamos de exponer, se aclaran bastante si se profundiza en el sentido de la MR. La condición A' significa que, en el marco de una estrategia conjunta imparcial y optimizadora, para un agente será racional maximizar su utilidad, dentro del límite de las *utilidades* que la estrategia asigna a los demás. La diferencia con la condición A es evidente, pues ella establecía la racionalidad de maximizar la propia utilidad, dadas las *estrategias* esperadas de los demás. El nuevo modelo de racionalidad implica la disposición a considerar, no las estrategias o acciones de los demás, sino sus utilidades esperadas en el marco de una estrategia conjunta, y actuar en consecuencia.

Así, se puede decir que "un maximizador *directo* es una persona que

[101] Sobre este argumento, las palabras más claras de Gauthier se encuentran en "Rational Constraint: Some Last Word", cit., p. 327. Reproducimos por extenso su conclusión: "Algunos defenderían que es racional actuar de un modo mutuamente ventajoso en situaciones de Dilema del Prisionero, siempre que los demás hagan lo mismo, porque uno gana más por su restricción de lo que pierde por la restricción propia. Este es un mal argumento que rechazo totalmente. Sin embargo, al centrarme en la restricción mutua en *MA*, mi rechazo tal vez quedó menos claro de lo que debería haber quedado. Es racional actuar de un modo mutuamente ventajoso en las situaciones de Dilema del Prisionero si uno gana más por su disposición a la restricción de lo que perdería por ejercitar de hecho la restricción. Éste es un buen argumento; es el argumento de *MA*; y no tiene nada que ver con el beneficio mutuo."

trata de maximizar su utilidad dadas las estrategias de aquellos con quienes interactúa. Mientras un maximizador *restringido* es una persona que, en algunas situaciones, trata de maximizar su utilidad dadas, no las estrategias, sino las utilidades de aquellos con quienes interactúa"[102].

Debe quedar claro, también, que la MR incluye a la MD como parte suya. La condición A' tiene dos partes: si el agente estima que un número crítico de otros agentes seguirán una estrategia conjunta, él mismo la seguirá. Si, por el contrario, cree que las posibilidades de la cooperación no son suficientes (por que se halla entre maximizadores directos, por ejemplo), entonces seguirá una estrategia directamente maximizadora. De este modo, lo que diferencia a un agente que actúa conforme a la condición A de un agente que lo hiciera conforme a la condición A' será exclusivamente la diferencia que resulte de sus distintas disposición *en las situaciones en que la cooperación es posible*.

Por último, queremos recalcar una vez más el dato, frecuentemente pasado por alto —como muestran los malentendidos a que nos hemos referido—, de que tanto la MR como la MD son "disposiciones" o, con las palabras de Julia Barragán, "conceptos de racionalidad"[103]. No se trata de "estrategias a largo plazo", ni de "meta-estrategias", ni de "compromisos con uno mismo"[104]. La relevancia de la propuesta de Gauthier radica precisamente en que, de lo que se trata es de transformar la racionalidad misma desde dentro; dando lugar a un "nuevo concepto de racionalidad". La cooperación exige, por tanto, adoptar ese nuevo concepto de racionalidad de acuerdo con

---

[102] Gauthier, D., *MA*, p. 167.

[103] Cfr. Barragán, J., "Las reglas de la cooperación", cit., p. 342 y ss. El artículo de la profesora Barragán es de los pocos que capta con claridad el carácter de las "disposiciones". Se puede considerar una excepción entre los teóricos bayesianos "ortodoxos".

[104] Se ha llegado incluso a comparar las "disposiciones" con los programas rivales que Axelrod enfrentó en el conocido concurso entre computadoras narrado en *La evolución de la cooperación*. Esta referencia nos permite una comparación iluminadora. Porque la conclusión general del estudio de Axelrod puede resumirse *grosso modo* en la tesis de que 'en términos de auto-interés, existen razones para *actuar* moralmente', mientras que lo que Gauthier quiere demostrar es que 'en términos de auto-interés, existen razones para *dejar de ser un agente auto-interesado* y convertirse en otro tipo de agente: un agente moral'.

el cual es individualmente racional cumplir los acuerdos optimizadores imparciales.

b) El argumento en favor de la maximización restringida.-

Si pusiéramos las exigencias de la racionalidad individual maximizadora y las exigencias de la cooperación en un mismo plano ¿Hacia cuál de ellas se inclinaría un agente racional? De otro modo, si en vez de decidir sobre la mejor estrategia a seguir en una situación de interacción, un agente racional hubiera de decidir sobre su propia disposición racional, ¿cuál de ellas adoptaría?

Básicamente este es el inicio del argumento de Gauthier en favor de la MR. El argumento se apoya, por tanto en dos presunciones: primero, que los agentes racionales son capaces de reflexionar sobre el modo de su interacción y modificar su disposición racional; segundo, que el único criterio que emplean en sus decisiones (sean estas entre estrategias o entre disposiciones) es la maximización de la utilidad esperada.

La segunda presunción se deriva directamente de la definición de racionalidad. La negación de la primera supondría afirmar, hobbesianamente, que somos "máquinas de maximizar a corto plazo". Ambas presunciones pueden considerarse, por tanto, plausibles.

Una tercera clave del argumento es la idea de que una disposición racional debe tener la capacidad de auto-sustentarse. Esto es, que, planteada una elección entre disposiciones, debería ser racional elegir una disposición dada sobre la base de los criterios que esa disposición propugna. En nuestro caso, parece que debería ser racional, en términos de maximización de utilidad, elegir ser un maximizador de utilidad, y no otro tipo de agente. Como es fácil imaginar, Gauthier sostiene que esto no ocurre en el caso de la maximización directa. Un maximizador directo, ante la alternativa de elegir la MD o la MR, elegiría la segunda *por motivos auto-interesados*, es decir, de acuerdo con los criterios de la MD. Por tanto, Gauthier afirmará que la MD no se sustenta a sí misma como disposición racional.

Pero todavía no se han demostrado suficientemente estas afirmaciones. Para hacerlo hay que mostrar simplemente que:

> "Dadas ciertas condiciones plausibles y deseables, un maximizador racional de la utilidad, enfrentado a la elección entre no aceptar ninguna restricción sobre sus elecciones en interacción y aceptar las restricciones requeridas por la concesión relativa *minimax*, elige esto último. Toma una decisión sobre cómo tomar las decisiones futuras; elige, sobre la base de la maximización de utilidad, no elegir nunca más sobre esa base."[105]

Esta demostración se realiza colocándonos en el lugar de ese agente que ha de optar entre disposiciones racionales y siguiendo sus argumentos. Se trata de un problema habitual de elección racional.

Se trata de determinar la utilidad esperada de elegir una u otra disposición. Como ya dijimos arriba, hay un número de situaciones en que el resultado de ambas disposiciones es idéntico, pues la MR es equivalente a la MD en las interacciones en que la cooperación no es posible. La diferencia entre disposiciones se reducirá a la diferencia de utilidad derivada de una y otra en las situaciones en que existe la perspectiva de la cooperación. Estas situaciones se distinguen porque (a) ofrecen la posibilidad de realizar un beneficio cooperativo, y (b) la matriz de pagos de la situación refleja que el equilibrio cooperativo es inestable, es decir, la desviación individual de la estrategia conjunta es individualmente más beneficiosa[106].

En este tipo de situaciones relevantes, cada agente puede calcular la utilidad derivada de adoptar una u otra disposición y compararlas. La disposición que permita mayor beneficio, será la elegida. Salta a la vista que el mejor pago en estos casos es el producto de defraudar mientras los demás se adecuan a la estrategia conjunta; el segundo lugar lo ocupa el resultado

---

[105] Gauthier, D., *MA*, p. 158.

[106] Si esto no fuera así, nos encontraríamos en una situación de coordinación. Cfr. arriba, punto 2.c) de este mismo capítulo.

cooperativo (todos cumplen), y el tercero el resultado derivado de la mutua no-cooperación (todos incumplen)[107].

Convencionalmente asignaremos el valor u al resultado no cooperativo, u' al resultado de la cooperación y u'' al beneficio del incumplimiento unilateral (beneficio del "gorrón"). Entonces, es evidente que la relación entre estas utilidades es $u'' > u' > u$. Pues bien, Gauthier sostiene que, dada esta relación entre los valores de utilidad, los agentes pueden plantearse dos argumentos contrarios sobre la cuestión de qué disposición adoptar[108]:

> *Argumento 1*: Supongamos que adopto la MD. Entonces, si espero que el otro agente base su acción en una estrategia conjunta, yo le defraudaré y mi utilidad esperada será u''. Si espero que actúe conforme a una estrategia individual, yo haré lo mismo y esperaré una utilidad u. Si la probabilidad de que el otro agente siga una estrategia conjunta es p, entonces mi utilidad global esperada será $[pu''+(1-p)u]$.
> Supongamos que adopto la MR. Entonces, si preveo que el otro agente va a seguir una estrategia conjunta, yo haré lo mismo y esperaré una utilidad u'. Si espero que emplee una estrategia individual, también la emplearé yo, consiguiendo u. De ese modo, mi utilidad global esperada es $[pu'+(1-p)u]$.
> Como u'' es mayor que u', $[pu''+(1-p)u]$ es mayor que $[pu'+(1-p)u]$ para cualquier valor de p distinto de 0 (y para $p=0$, ambos son iguales). De donde se sigue que, para maximizar mi utilidad global esperada, debo adoptar la maximización directa (MD).

> *Argumento 2*: Supongamos que adopto la MD. Entonces debo esperar que los demás empleen estrategias maximizadoras individuales cuando

---

[107] Es evidente que aún hay un resultado peor: aquél en que uno coopera y todos los demás defraudan, explotando su ingenuidad. Pero Gauthier no considera por el momento este caso, dando por supuesto que ningún agente cometería un error de apreciación tan grave.

[108] Transcribimos los argumentos tal como son expuestos en *MA*, pp. 171-172; pero empleamos también algunos comentarios incorporados en la versión de Edward McClennen en "Constrained Maximization and Resolute Choice", cit., pp. 98-101.

interactúen conmigo; yo haré lo mismo y recibiré una utilidad *u*.
Supongamos que adopto la MR. Entonces, si los demás están condicionalmente dispuestos a la MR, puedo esperar que sigan una estrategia conjunta cooperativa cuando interactúen conmigo; yo haré lo mismo, y recibiré una utilidad *u'*. Si no muestran esa disposición, yo emplearé una estrategia maximizadora y esperaré una utilidad *u*. Si la probabilidad de que un agente esté dispuesto a la MR es *p*, entonces mi utilidad global esperada es $[pu'+(1-p)u]$.
Dado que *u'* es mayor que *u*, $[pu'+(1-p)u]$ es mayor que *u* para cualquier valor de *p* mayor que 0 (y para $p=0$, ambos son iguales). De donde se sigue que, para maximizar mi utilidad global esperada, debo adoptar la maximización restringida (MR)[109].

Evidentemente, ambos argumentos no pueden ser correctos en el mismo sentido, puesto que ofrecen conclusiones opuestas. El primero está basado en el concepto de estrategia dominante, es decir, en la lógica típica de las situaciones tipo Dilema del Prisionero: si el adversario elige maximizar, hago mejor maximizando (evito ser explotado); si el adversario elige cooperar,

[109] Un análisis muy iluminador de este argumento puede verse en Holly Smith, "Deriving Morality from Rationality", en Vallentyne, P. (ed.), *Contractarianism and Rational Choice*, cit., pp. 229-253; p. 234. Para un desarrollo más abstracto y crítico, cfr. Maarten Franssen, "Constrained Maximization Reconsidered: An Elaboration and Critique of Gauthier's Modelling of Rational Cooperation in a Single Prisoner's Dilemma", *Synthese*, 101 (noviembre, 1994), pp. 249-272. Por otro lado, Gauthier mismo ofrece una versión "intuitiva" del mismo cuando, en 1978, escribe: "Supongamos que soy un hombre económico racional. Me doy cuenta de que, si concierto un acuerdo para lograr un resultado óptimo y así superar las ineficiencias externas [del mercado], lo romperé siempre que pueda beneficiarme con ello. Me doy cuenta también de que tu lo sabes [...]. Por tanto, sé que seré excluido de los acuerdos voluntarios con socios racionales e informados. Pero esto es desventajoso para mí. Si yo considerara racional conformarme a esos acuerdos (supuesto que los demás lo hicieran), entonces podría formar parte de acuerdos voluntarios mutuamente beneficiosos con otros agentes que también consideraran racional la conformidad. Luego, como hombre económico racional que trata de maximizar su propia utilidad, debo considerar racional cambiar mi punto de vista, y modificar mi concepción de la racionalidad económica de modo que, cuando espere mayor utilidad si todos se adhieren a un acuerdo que la resultante de que todos lo violen, y espere además que todos se adhieran, entonces consideraré racional adherirme yo mismo, incluso aunque podría maximizar mi propia utilidad violándolo. Así modificada, mi concepción de la racionalidad puede denominarse maximización de la utilidad individual *restringida*." ("Economic Rationality and Moral Constraints", *Midwest Studies in Philosophy*, III, pp. 75-96; pp. 91-92).

también hago mejor maximizando (me beneficio explotándolo). Este argumento —dice Gauthier— sería válido si la probabilidad de que otros agentes actuasen cooperativamente fuese independiente de la propia disposición. Y ése no es el caso. Dado que la disposición a la cooperación es (según la condición A') condicional —un MR actuará cooperativamente sólo con aquellos que estén igualmente dispuestos—, un MD no tendrá las oportunidades de beneficiarse que sí se le presentan al MR. Si todos tuvieran las mismas oportunidades de entrar en empresas cooperativas (o acuerdos), entonces los MD obtendrían mayores beneficios, según establece el argumento de la estrategia dominante. Pero es precisamente ahí donde está la diferencia. El argumento segundo tiene en cuenta que sólo los MR serían admitidos como partes en las empresas cooperativas (pues son fiables), mientras que a los MD se les privaría de esta oportunidad. El argumento segundo resulta más plausible y, por tanto, su conclusión puede aceptarse como la decisión racional desde el punto de vista de la maximización de utilidad.

Sin embargo, la victoria de la maximización restringida no es, ni con mucho, tan sencilla. El argumento que la apoya contiene un pre-supuesto no defendido. En él, se ha dado por sentado que todos los agentes son conocidos por los demás *tal como son*. Es decir, el agente que razonaba, ha contado con una información perfecta sobre la disposición de sus adversarios (y ha supuesto también que su disposición era públicamente conocida). Si esta improbable condición se diera, el argumento sería completamente convincente. Esta condición, denominada por Gauthier "transparencia"[110], podría adscribirse por hipótesis a las personas idealmente racionales que protagonizan el contrato social y seguir adelante con la teoría. "Pero —prosigue Gauthier— queremos que nuestros supuestos ideales estén en relación con el mundo real. Si la MR vence a la MD sólo en el caso de que todas las personas sean transparentes, entonces no habremos mostrado la racionalidad de las restricciones morales bajo condiciones reales. Habremos refutado al *Tonto*, pero al precio de despojar

[110] Cfr. *MA*. pp. 173-174.

a nuestra refutación de cualquier significado práctico"[111].

Para devolver al argumento su significado práctico habría que sustituir la transparencia por otro supuesto más débil, que Gauthier denomina "translucidez". Si, como parece más plausible, las personas somos translúcidas, cada agente puede adivinar con cierto grado de probabilidad la disposición de sus eventuales socios; pero también cabe la posibilidad de ser engañado o engañar.

La translucidez obliga a introducir en el argumento un gran número de nuevas variables. Gauthier las reduce a tres:

- La ya empleada probabilidad *p* de que un MR encuentre a otro MR, se reconozcan mutuamente y cooperen con éxito.
- La probabilidad *q* de que un MR yerre en su apreciación (confunda a un MD con un MR) y sea explotado.
- La probabilidad *r* de hallar MRs, que dependerá únicamente de su proporción en la población total.

Aún se podrían introducir otras variables que complicarían (y generalizarían) el modelo, tales como la probabilidad del "mutuo error" de dos MR que les llevara a tratarse como dos MD y "desperdiciar" una ocasión para cooperar o la probabilidad de que un MR explote por error a otro MR al confundirlo con un MD, etc.[112]. No profundizaremos en este camino; sólo queremos dejar constancia de que, incluso el complejo argumento de la translucidez supone una simplificación.

Para calcular los pagos teniendo en cuenta las probabilidades *p*, *q* y *r*, hay que añadir, a las tres utilidades anteriores *u* (no cooperación), *u'* (cooperación) y *u''* (incumplimiento unilateral), la utilidad derivada de ser

---

[111] *MA*, p. 174.

[112] Maarten Franssen, en el art. cit. (p. 252), toma en cuenta estas posibilidades y formula un argumento generalizado (incluyendo la "explotación accidental", que es el pago "extra" que Gauthier obvia). En esta línea, y sólo como ejemplo de refinamiento, reproduzco un texto de H. Lottenbach ("Expected Utility and Constrained Maximization: Problems of Compatibility", *Erkenntnis*, 41, pp. 37-48; p. 42): "Nótese que del hecho de que dos MRs se encuentren no se sigue que cooperen. Si, por ejemplo, el primero cree que el segundo cree que él es un MD, no cooperará".

explotado, a la que daremos un valor 0.

Teniéndolo todo en cuenta se puede calcular la utilidad esperada, para MRs y MDs, en las situaciones que analizaba el argumento anterior: posibilidad de cooperación mutua, o posibilidad de defraudar logrando un beneficio mayor.

Un MR espera una utilidad u excepto cuando coopera satisfactoriamente con otros MRs —consigue u'— o cuando es defraudado por un MD —consigue 0. La probabilidad de que se dé el primer caso es igual a la probabilidad combinada de que interactúe con otro MR (proporción r), y que se reconozcan mutuamente (probabilidad p). La probabilidad combinada es rp. En este caso, se logra la utilidad u', con un beneficio neto de $(u'-u)$ sobre la utilidad "segura" de la no-cooperación. Así que el incremento de utilidad esperada que proporcionan los casos de interacción cooperativa a los MR vale $[rp(u'- u)]$. La probabilidad de ser defraudado es, por otra parte, la combinación de encontrarse con un MD $(1-r)$ y no reconocerlo (q), que equivale a $(1-r)q$. El pago es 0, de forma que se puede considerar que en estos casos el MR pierde la utilidad "segura" u de la no-cooperación. Tomando las dos posibilidades en cuenta, la utilidad global esperada por el MR es la siguiente: $\{u+[rp(u'-u)]-(1-r)qu\}$.

Por su lado, el MD espera siempre una utilidad u, excepto cuando defrauda a un MR (él mismo nunca puede ser defraudado, ya que nunca coopera). La probabilidad de que defraude equivale a la probabilidad combinada de que encuentre un MR (r) y lo reconozca sin ser reconocido (q), es decir rq. En este caso el incremento de utilidad que espera es $(u''-u)$ —la utilidad derivada de su incumplimiento menos la utilidad que de todas formas recibiría. El efecto de estos casos es un incremento de su utilidad esperada en un valor $[rq(u''-u)]$. Por tanto, la utilidad global esperada por un MD será $\{u+[rq(u''-u)]\}$.

La utilidad global de un MR, $\{u+[rp(u'-u)]-(1-r)qu\}$, será mayor que la de un MD, $\{u+[rq(u''-u)]\}$, dependiendo de que la razón entre la probabilidad p (cooperar satisfactoriamente) y la probabilidad q (ser explotado) sea más grande que la razón entre el beneficio de defraudar y el beneficio de la cooperación.

Gauthier argumenta que, suponiendo una distribución igualitaria de MRs y MDs, y suponiendo unas habilidades aceptables (incluso bastante escasas)

para detectar las disposiciones de sus semejantes, los MRs puede esperar mejores pagos que los MD en un mundo de personas translúcidas (de modo que se mantiene la validez del argumento en favor de la MR)[113]. Si la disposición MR tiene asociada una utilidad esperada mayor que la disposición MD, es de esperar que el número de MDs decrezca al irse convirtiendo progresivamente en MRs . Dicho de otra forma, unas personas con un grado no muy exigente de translucidez, encontrarían racional la moralidad.

En resumen, Gauthier mantiene que es racional, en términos de maximización de la utilidad, convertirse en MR. El argumento depende un supuesto bastante contingente, pero aceptablemente plausible, en la medida en que no es un supuesto demasiado exigente y refleja las condiciones del mundo real.

Es un argumento que apela a la racionalidad maximizadora, por lo tanto, está diseñado para convencer al *Tonto*. Éste, tiene ante sí la posibilidad de lograr sus objetivos en un grado determinado si mantiene su disposición MD; y de lograrlos en un grado mucho mayor si adopta la disposición a cooperar. Si es un agente racional perfecto que conoce esto (y es consciente de que su necesidad) ¿acaso dudará en convertirse en un agente cooperativo y fiable?, y al hacerlo encontrará racional cumplir los acuerdos pactados y podrá ser llamado justo —de acuerdo con la tercera ley de la naturaleza, según Hobbes.

No sabemos si el *Tonto* quedaría convencido por los argumentos de Gauthier, pero sí está claro que una cohorte de herederos y albaceas reivindican su patrimonio, reviviendo su objeción en las críticas dirigidas a esta parte de la teoría contractual de la moral.

[113] Cfr. *MA*, p. 177, para ver los datos concretos del supuesto en que basa Gauthier su afirmación.

c) Las objeciones de los herederos del *Tonto*.-

Esperamos que nadie se ofenda por nombrarle heredero de tan ilustre *Tonto*. El nombramiento se debe a que este personaje representa la racionalidad como maximización en estado químicamente puro; por tanto, sí se le adivina un parentesco con los defensores de la Teoría Bayesiana de la Decisión que, de modo predominante, ocuparán este epígrafe. Pues, en efecto, dado que Gauthier pretende establecer su defensa de la MR sobre una decisión paramétrica, la mayor parte de las críticas provienen de especialistas en este campo, que recusan alguno de los pasos o los supuestos del razonamiento expuesto arriba.

En dos palabras, la objeción actual reproduciría la clásica: Por muy plausible que parezca el argumento, las demandas de la justicia (si la justicia ha de ser entendida como el cumplimiento exclusivo de los acuerdos imparciales) no corren paralelas a las de la racionalidad. Bien porque no está claro que sea beneficioso convertirse en un MR, bien porque aun suponiendo que cabe esa transformación, quepa dudar de su estabilidad en el tiempo, bien porque aunque se haya demostrado la racionalidad de ser un MR entre MRs, nada se ha dicho del caso en que todavía nadie ha dado el primer paso hacia la moralidad. Unos y otros razonamientos se suceden para mostrar que Gauthier no ha avanzado mucho respecto a Hobbes. A pesar de todo, un agente racional sigue considerando más beneficioso defraudar.

Quienes sustentan estas posiciones son mayoritariamente, como quedó dicho, teóricos de juegos fieles a la teoría de la utilidad. Intentaremos una revisión de algunos de sus argumentos —obviamente sin ánimo de exhaustividad.

El planteamiento mismo de Gauthier es blanco de las primeras críticas. Lo que Gauthier plantea es una "competición" entre la MR y la MD para determinar cuál de estas disposiciones ofrece una mayor utilidad esperada. Pues bien, autores como Hans Lottenbach[114], Govert Den Hartogh[115] y Julian

[114] "Expected Utility and Constrained Maximization: Problems of Compatibity", *Erkenntnis*, 41 (1994), pp. 37-48.

[115] "The Rationality of Conditional Cooperation", *Erkenntnis*, 38 (1993), pp. 405-427.

Nida-Rümelin[116] han negado que tal planteamiento tenga sentido dentro de la Teoría de la Decisión. Lottenbach expresamente se propone "mostrar que incluso aunque la noción de elegir una disposición se hiciera inteligible, la MR (o cualquier otra propuesta de disposición maximizadora) no puede ser defendida por la teoría de la utilidad esperada"[117]. El núcleo de su crítica se cifra en la disimetría entre MR y MD. Según Gauthier propone, habría que comparar (desde una concepción *estándar* de la racionalidad como maximización) la utilidad esperada derivada de la elección de cada disposición. Pero comparar utilidades en este sentido exige que sea posible representar esas utilidades conforme a los axiomas de Von Neuman-Morgenstern (esto es, conforme a los axiomas de la teoría de la utilidad aceptada). Lo cual no presenta problemas para la MD, pero, sí los presenta para la MR. Por definición, la MR no representa un comportamiento ajustable a los axiomas de la teoría clásica de la utilidad. Gauthier compara en sus argumentos utilidades esperadas (las derivadas de las acciones de uno y otro agente); pero —sostiene Lottenbach— los MRs *no tienen* utilidades esperadas (en el sentido von Neuman-Morgenstern). Por otro lado, ni siquiera las utilidades que Gauthier adscribe a los MDs estarían correctamente estimadas. Porque la versión heredada de la maximización trata de la utilidad esperada de las acciones, no de las disposiciones (que es a lo que Gauthier la aplica). Lottenbach cree que la utilidad esperada de un "maximizador directo entendido como un seguidor de la disposición MD" debería distinguirse de la utilidad esperada de un maximizador directo simplemente entendido conforme a la teoría de la utilidad heredada. En su opinión, la teoría bayesiana no sirve para comparar disposiciones[118]. Habría que emplear otra teoría de la utilidad según la cual —argumenta— la elección entre MR y MD tendría —para un maximizador de la utilidad— el mismo carácter dilemático (estructura Dilema del Prisionero) que tiene la elección de acciones desde el punto de vista de la teoría clásica de la

[116] "Practical Reason, Collective Rationality and Contractarianism", en Gauthier y Sugden (eds.), *Rationality, Justice and the Social Contract*, cit., pp. 53-74.

[117] Lottenbach, H., art. cit., p. 37.

[118] Cfr. Lottenbach, art. cit., p. 42, para un ejemplo en defensa de esta tesis.

utilidad. Y, dada esta estructura, lo racional es no-cooperar.

Una consecuencia ulterior de esta crítica es que invalida el argumento de Gauthier en favor del cumplimiento. Este argumento sostiene que si adoptar una disposición (la MR, para nuestro caso) maximiza la utilidad, entonces es racional elegir siempre conforme a la regla de la disposición. Pero —como muy agudamente observa Lottenbach— tal argumento no convencerá a quien, como él mismo, no identifique la racionalidad práctica con la maximización de utilidad en el nivel de las disposiciones.

Como vemos, el argumento de Lottenbach es implacable con la tesis de Gauthier. En este punto, Lottenbach representa un fantasma que acompaña invariablemente todo intento de avanzar en la comprensión de la conducta racional. Se entiende que la réplica de Gauthier simplemente consista en poner de manifiesto, una vez más, la incoherencia de la *posición ortodoxa* que, en este caso, representa Lottenbach:

> "Suponer, como hace la teoría ortodoxa, que la imposición de tal restricción es imposible, es ver la racionalidad en algunos aspectos como un obstáculo para maximizar la propia utilidad, en vez de como un instrumento para ello. La teoría ortodoxa trata la racionalidad como auto-frustrante en situaciones con estructura de Dilema del Prisionero. Mi comprensión alternativa de la capacidad racional elimina esta incoherencia."[119]

Sin embargo, los contumaces defensores de la ortodoxia son inasequibles al desaliento, como demuestran las conclusiones de Lottenbach: "*Ninguna* teoría de la utilidad del tipo von Neuman-Morgenstern puede ser coherente sin excluir la cooperación en los Dilemas del Prisionero. Aplicar la teoría no a las decisiones particulares, sino a las disposiciones o planes, no elimina ninguna incoherencia. La teoría de la utilidad de las disposiciones o planes no supone un avance sobre la teoría de la utilidad para elecciones particulares *en los*

[119] Gauthier, "Economic Man and the Rational Reasoner", en Nichols y Wright (eds.), *From Political Economy to Economics - And Back?*, San Francisco ICS Press, 1990, pp. 105-131; p. 122.

términos de esta última. Además, al nivel de las decisiones sobre disposiciones o planes, los maximizadores de utilidad se encontrarían de nuevo con un Dilema del Prisionero, y su teoría prescribirá la no-cooperación"[120].

Den Hartogh y Nida-Rümelin ofrecen otra razón en contra del planteamiento de Gauthier. Aunque formulado de modos muy diferentes —muy técnico el primero, más intuitivo el segundo—, creemos que el argumento es en el fondo el mismo. La tesis principal se resume en que "es un atributo esencial de una persona racional el ser relativamente libre de sus determinaciones disposicionales"[121]. Si esta tesis es correcta, entonces no hay posibilidad de elegir entre disposiciones o, mejor dicho, tal elección sería irrelevante para las decisiones futuras.

Empleando el lenguaje de Nida-Rümelin, podemos decir que los argumentos expuestos vienen a incidir en que no es posible re-construir la racionalidad colectiva a partir de una concepción simplemente maximizadora de la racionalidad individual. Ni siquiera un cambio individual en las preferencias aseguraría la resolución del conflicto entre racionalidad individual y colectiva. Una elección de disposición basada en la maximización —si acaso fuese pensable desde la concepción ortodoxa— produciría un resultado siempre inestable y no siempre maximizador[122].

Esto nos lleva a otro grupo de argumentos que, implícita o explícitamente, aceptarían el marco de discusión establecido por Gauthier. Den Hartogh nos proporciona una primera crítica de este grupo, con los que podemos denominar "argumento de la simetría" y "dilema del contribuyente II".

El argumento de la simetría parte de dos premisas. La primera es que la disposición cooperativa es condicional: un agente está dispuesto a cooperar *si* espera que aquél con quien interactúa esté igualmente dispuesto. La segunda

---

[120] Art. cit., p. 45.

[121] Nida-Rümelin, J., art. cit., p. 56. Den Hartogh lo expresaría diciendo que en una decisión estratégica no se puede suponer que los parámetros en los que descansa la deliberación estén dados.

[122] Cfr. Nida-Rümelin, J., art. cit., p. 71.

premisa es que en un par de disposiciones asimétricas, al menos una no es racional. De donde se deduce que es imposible formar intenciones racionales simétricas de cooperar condicionalmente. Pues cada agente cooperaría sólo si espera que otro lo haga, y si éste muestra una disposición semejante, entonces ninguno de los dos cooperará, ya que nunca se dará la condición de su disposición[123]. La disposición MR sólo se adoptará si se supone cierta asimetría entre los agentes: "La mutua cooperación, tal como la formula Gauthier, sería parcialmente asimétrica. Un elector cooperativo está preparado a elegir la cooperación si y sólo si espera que su compañero elija asimismo cooperar; y esta condición se satisface si elige cooperar *incondicionalmente*"[124]. De estas dos disposiciones asimétricas —cooperación incondicional y cooperación condicional— al menos una debe ser irracional. Pues bien, aunque parezca paradójico, es la cooperación condicional la que resulta insostenible (lo que no quiere decir que la cooperación incondicional sea menos irracional). Veamos por qué: Una persona que respondiera no-cooperativamente a los cooperadores incondicionales y en todos los demás casos actuara como un cooperador condicional obtendría mejores pagos que el cooperador condicional (obtendría sus mismos resultados más el beneficio de explotar a los cooperadores incondicionales).Por lo tanto, sería racional para los cooperadores condicionales cambiar su disposición por esta nueva "cooperación condicional pero explotadora" (al menos en la medida en que existan cooperadores incondicionales en la población).

El argumento de la simetría quiere poner de relieve que ningún esquema cooperativo que exija disposiciones asimétricas será estable. Pero *formar* disposiciones condicionales simétricas parece imposible. subrayamos *formar*, porque *tener* disposiciones simétricas sí es posible. De hecho, las tenemos muy

---

[123] Obsérvese que este círculo se asemeja al que Gauthier mismo plantea al analizar las condiciones de la decisión estratégica: La decisión (o estrategia) de un agente ha de ser una respuesta racional a las acciones que espera que otros realicen; pero las acciones de éstos también deben ser respuesta racional a las acciones del agente etc. Sin embargo, este círculo vicioso se resuelve asignando probabilidades a las acciones posibles de los demás y seleccionando una estrategia maximizadora (que puede ser una lotería entre estrategias puras). Den Hartogh obvia este camino para una posible solución.

[124] Den Hartogh, G., art. cit., p. 411.

a menudo (siempre que realizamos actividades cooperativas o coordinadas). Por eso, la conclusión de Den Hartogh es que estas disposiciones no pueden explicarse a partir de conciencia que los agentes tienen de las disposiciones de los demás, sino a partir de una fuente independiente: una razón que nos movería en ausencia de razones condicionales.

El segundo argumento ha sido calificado como "el dilema del contribuyente II" porque, como aquél dilema, se basa en cuestionar la racionalidad de la cooperación cuando el número de cooperadores es tan grande que la importancia relativa de cada acción cooperativa individual es mínima. El argumento es tan razonable como el del dilema original. La transcripción de un breve párrafo bastará para mostrar su plausibilidad: "Debemos estar dispuestos a cooperar en la producción de bienes públicos con quienes estén similarmente dispuestos. ¿Por qué? ¿Porque en otro caso los productores cooperativos de bienes públicos nos excluirían de los beneficios derivados de ellos? Esto no tiene sentido. Si es posible excluir a la gente del beneficio derivado de un bien, es que no es un bien público..."[125].

Desde luego el argumento de Gauthier no es tan simple, pues "excluir" se refiere, en su contexto, a eliminar las oportunidades ulteriores de cooperar. Aún así, Den Hartogh lo analiza como si se tratara de una cuestión de producción de bienes públicos habitual. Y su resultado es que sólo en el caso de que la decisión de un agente resultara ser decisiva para la producción del bien público (porque fuese, por así decir, el "voto" que falta para la mayoría absoluta), sería racional para él adoptar la disposición cooperativa. En los demás casos, si el bien público va a ser producido de todos modos, es más beneficioso defraudar (o haber suscrito hipócritamente el acuerdo para la producción de la cooperación). Si no se va a producir de ningún modo, nada cambia por el hecho de haber sido insincero[126].

No hay duda de que los argumentos de Den Hartogh —como los comentados anteriormente— son atractivos y convincentes. Pero lo son en un

[125] Den Hartogh, G., art. cit., p. 418.

[126] Cfr. Den Hartogh, G., art. cit., p. 419-420.

sentido que no hay que perder de vista. Están formulados desde dentro de un paradigma teórico (la Teoría Bayesiana de la Decisión) relativamente cerrado. Resultan convincentes precisamente porque hemos adoptado —tanto Gauthier en su moral por acuerdo como nosotros en este trabajo— ese punto de vista sobre la racionalidad. Pero su estatuto en relación con una teoría moral contractual es precario. La teoría moral se formula desde el marco de la Teoría de la Decisión, pero para transcenderlo o, dicho de otra forma, para mostrar justamente el tipo de desarrollo que estos argumentos ocultan.

Creemos que Julia Barragán representa y sintetiza muy bien el enfoque bayesiano ortodoxo cuando, refiriéndose a la defensa gauthieriana de la MR, dice que "a la hora de plantearse la estabilidad de la solución, necesita apelar a una modificación de la matriz de pagos"[127]. Porque, desde el punto de vista de la teoría de la utilidad, la matriz de pagos asociada a la estructura Dilema del Prisionero implica *siempre* que la solución cooperativa es inestable. Estos autores llevan hasta sus últimas consecuencias el compromiso normativo con la maximización de la utilidad esperada, de forma que constantemente interpretan los argumentos en términos de estrategias y utilidades. Como mucho, introducen conceptos que podrían considerarse afines a las disposiciones, como las meta-estrategias, o los planes[128]. Su enfoque es opaco al argumento de Gauthier.

Si el intento superador de Gauthier fuese único en su género, la persistencia de este tipo de críticas quizá invitaría a un replanteamiento. Pero no es así. Otros teóricos mucho menos sospechosos que Gauthier han ensayado "salidas" no ortodoxas al Dilema del Prisionero. Algunas de ellas serán

[127] Barragán, J., art. cit., p. 351. Según este texto, la modificación introducida en el argumento segundo de Gauthier (cfr. epígrafe anterior) sobre las oportunidades de los MRs, supondría una variación en la matriz de pagos respecto al argumento primero. Si entendemos bien la posición de Barragán, en el segundo argumento la MR es preferida porque la situación-ejemplar para la que se elige ya no es más un Dilema del Prisionero.

[128] De hecho, sólo Julia Barragán habla decididamente de una "nueva racionalidad"; pero finalmente concluye, como hemos visto, que la salida cooperativa sigue siendo inestable: juzga que Gauthier no va lo bastante lejos y que no logra escapar de las críticas "ortodoxas".

aludidas más abajo[129].

Además, el hecho de que los bayesianos ortodoxos se hayan centrado en el mecanismo —una elección paramétrica entre disposiciones racionales— no quiere decir que el peso del argumento de Gauthier esté en esa elección (que muy bien pudiera ser criticable) sino en la necesidad racional que ella representa:

> "Hemos defendido la racionalidad de la MR como una disposición para elegir mostrando que sería racionalmente elegida. Ahora bien, este argumento no es circular; la MR es una disposición para la elección estratégica que sería elegida paramétricamente. Pero la idea de una elección entre disposiciones para elegir es un mecanismo heurístico para expresar la necesidad subyacente de que una disposición racional sea maximizadora de la utilidad. En los contextos paramétricos, la disposición a realizar elecciones directamente maximizadoras es incontrovertiblemente maximizadora de la utilidad. Podemos, por tanto, emplear el mecanismo de una elección paramétrica entre disposiciones para mostrar que en los contextos estratégicos la disposición a realizar elecciones restringidas, y no elecciones directamente maximizadoras, maximiza la utilidad. Sin embargo, debemos enfatizar que la clave de nuestro argumento no es la elección misma, sino el carácter maximizador de la disposición (en virtud del cual es digna de ser elegida)."[130]

Sobre el carácter maximizador de la disposición a la MR se pueden verter pocas dudas. Se puede observar que las críticas que hemos analizado se dirigen contra el proceso de elección incorporado en el argumento de Gauthier,

---

[129] Sobre estas propuestas —especialmente sobre las que no podremos recoger con detalle en este trabajo— puede verse Yanis Varoufakis, "Modern and Postmodern Challenges to Game Theory", *Erkenntnis*, 38 (1993), pp. 371-404.

[130] Gauthier, D., *MA*, p. 183.

pero pocos autores (por muy ortodoxos que sean) negarían que la cooperación se produce justamente porque es beneficiosa. Algunos de ellos (por ejemplo Nida-Rümelin) incluso reivindican un tipo de restricción más fuerte que la representada por la MR, porque creen que es el único modo de escapar a la incoherencia colectiva de la racionalidad como maximización.

Podemos concluir, por tanto, que si acaso el mecanismo heurístico de "elegir una disposición"[131] vale como argumento en favor de alguna disposición racional para la interacción estratégica, la disposición favorecida será la MR; precisamente la disposición que hace racional el cumplimiento de los acuerdos pactados.

d) Transparencia y translucidez.-

Si el planteamiento del argumento de Gauthier ha suscitado críticas como las que acabamos de ver, su supuesto de la translucidez —que por sí mismo lo debilita considerablemente— no ha quedado atrás en este orden. Con la agravante de que, si se puede decir que el argumento de Gauthier, tal como está planteado (como una renovación de la concepción de la racionalidad), es resistente a las críticas provenientes de los teóricos de juegos "ortodoxos", resulta por el contrario muy sensible a las provenientes de quienes aceptan la línea principal del mismo, y rechazan algunos de sus elementos o sus conclusiones. Entre estos últimos, predominan los que cuestionan el poco riguroso expediente de la "translucidez".

Ciertamente, en este punto hay que descartar un gran número de críticas que podemos calificar de "triviales". Muchas de ellas simplemente denuncian que la hipótesis de la translucidez no es verosímil, que se trata de una hipótesis

[131] Sobre la posibilidad de elegir una disposición volveremos más abajo. Pero, anticipando las posibles críticas hacia esa posibilidad, hay que notar que, como destaca Park (*op. cit.*, p. 155) "a lo largo de la historia de la filosofía, los filósofos morales han sugerido varias formas de soluciones disposicionales. Piénsese en Aristóteles, Hobbes, Dewey, Rawls, MacIntyre y Williams."

ad hoc, etc.

Por supuesto que se trata de una hipótesis *ad hoc* —que se podría haber obviado introduciendo una hipótesis aún menos plausible, como la transparencia. Ahora bien, es una hipótesis que, paradójicamente, dificulta el argumento, en vez de facilitarlo, de manera que no entendemos que sea criticable en este concepto. Ya explicamos arriba el motivo de su empleo: se trata de hacer verosímil el razonamiento en favor de la MR, acercándolo a unas condiciones más reales. En cuanto a que el supuesto sea más o menos ajustado a la "realidad", es una materia opinable. También comentamos que la precisión del "grado de transparencia" de los agentes tiene una importancia relativa, pues Gauthier muestra que unas condiciones de translucidez poco exigentes son suficientes para validar el argumento. Es decir, sería difícil imaginar las condiciones bajo las cuales su argumento no funcionase. Habría que suponer agentes extraordinariamente opacos e hipócritas, o excepcionalmente ingenuos e incapaces de adivinar las intenciones de sus semejantes. Tal vez la opacidad pueda considerarse una característica deseable de los perfectos maximizadores de utilidad que pueblan nuestro estado natural, pero no así la ingenuidad.

La idea de que entre maximizadores sería racional la opacidad y la hipocresía (en la medida en que llevasen a alcanzar el objetivo del agente) presta su base a la crítica de la hipótesis de la translucidez mejor construida —que tomamos como representativa. Se trata de la desarrollada por Geoffrey Sayre-McCord[132].

Sayre-McCord analiza perspicuamente la importancia de la translucidez en el argumento de Gauthier. Ofrece incluso una detallada tabla de los valores que podrían adoptar las probabilidades *p*, *q* y *r* para que el pago de ser moral igualase al pago de ser un "egoísta ilustrado" (alguien que finge ser moral pero que se beneficia de explotar a sus socios cuando tiene ocasión). Por ejemplo, si en la población hay un 75% de agentes morales (MR en los términos de Gauthier), un 70% de probabilidad de ser correctamente identificado y de identificar a los demás será suficiente para considerar racional adoptar la

[132] "Deception and Reasons to Be Moral", en Vallentyne, P. (ed.), *Contractarianism and Rational Choice*, cit., pp. 181-195.

disposición moral; sin embargo, si el porcentaje de agentes morales es sólo del 33%, una persona racional habrá de estimar en un 80% la probabilidad de ser correctamente identificado e identificar, para optar ella misma por la moralidad.

Hay dos buenos argumentos en favor de la translucidez (y de su papel en la defensa de la MR): el primero es que la posibilidad de identificar correctamente al prójimo es de un 50% con sólo recurrir al azar; se puede aceptar, por tanto, que todo el mundo tiene, al menos, una probabilidad del 50% de ser correctamente identificado. Si a esto se suma la habilidad y la experiencia, es plausible contar un alto grado de certeza en este aspecto, lo que favorece a la MR. El segundo argumento es que la opacidad es arriesgada. Un agente opaco tiende a ser excluido de la cooperación independientemente de su verdadero carácter, sólo por el riesgo asociado a interactuar con él. A la vista de esto, las personas opacas tendrán razones para incrementar su translucidez, para poder optar a la cooperación[133].

Lamentablemente, estos dos argumentos resultan menos convincentes si se tiene en cuenta que —como dice Sayre-McCord— la distancia entre lo que la gente parece y lo que realmente es puede ser muy grande. Respecto al argumento de la probabilidad de acertar la disposición del otro, Sayre-McCord recuerda que no es normal recurrir al azar, sino que las personas preferirán basar sus estimaciones en las evidencias fiables. Y, lógicamente, los defraudadores tendrán buen cuidado de mostrar las evidencias que califican a los cooperadores. Si tienen éxito, entonces los demás les juzgarán erróneamente la mayoría de las veces. Sayre-McCord sugiere que habría incluso un tipo de agente (el "mega-opaco" o "trans-opaco") que tendría tal éxito en sus engaños que la probabilidad de ser identificado correctamente tendería a cero (muy lejos del 50% con que contaba el primer argumento)[134].

En cuanto al segundo argumento, funciona para agentes opacos que no puedan cambiar su apariencia de opacidad pero, de nuevo, resulta irrelevante para el egoísta ilustrado transopaco.

No es necesario añadir que se podrían aducir muchos ejemplos prácticos

---

[133] Cfr. Sayre-McCord, G., art. cit., p. 191.

[134] Sobre esta sugerente idea, también puede verse Franssen, M., art. cit., p. 266 y ss.

que confirmarían hasta qué punto la tesis de Sayre-McCord es empíricamente correcta. Su idea clave es que la habilidad del egoísta racional es mucho mayor que la que Gauthier supone y que, en cuanto el egoísta detecte que la mejor estrategia podría no ser la moralidad, sino un refinado fingimiento de la misma, lo adoptará con todas sus consecuencias.

Con todo, Sayre-McCord admite que, entre quienes ya han adoptado la MR sería racional incrementar su translucidez, para reconocerse entre ellos y cooperar en más ocasiones (pese a que esto les ocasionara pérdidas al ser explotados también con más frecuencia). Pero entre quienes pueden elegir no sólo su carácter, sino también su apariencia (entre "buenos actores", diríamos) lo más racional es elegir ser egoístas ilustrados transopacos. Para algunos, el coste del engaño constante (que implica también cierto auto-engaño[135]) no compensará el beneficio esperado de la explotación de otros. Pero, en todo caso, se tratará de un cálculo que dependerá de cómo evalúe cada uno la utilidad de defraudar a sus semejantes y el coste relativo de la perfecta hipocresía.

Pese a su conclusión ambivalente, este argumento es bastante nocivo para la MR, puesto que, una vez más, pone de manifiesto que tal vez sea racional ser moral, pero no lo es necesariamente *volverse* moral. La doble vertiente de la conclusión de Sayre-McCord parece permitir una solución "comunitaria": puesto que la MR es estable entre quienes ya están dispuestos a ella y *todos* prefieren una sociedad repleta de MRs (aunque *cada* egoísta transopaco particular prefiere seguir siéndolo), es plausible que las instituciones socializadoras se diseñasen de modo que el agente medio tendiera a considerar muy costosa (en términos personales) una vida de engaño y fraude hacia sus semejantes. Mas se trata de una solución aparente, que dependerá en última instancia de las motivaciones de los miembros de la sociedad y del coste relativo que para la comunidad tenga la presencia de agentes transopacos. En una hipótesis de perfecta racionalidad, quizá la socialización de los nuevos miembros les dispondría a la MR, pero la aplicación práctica de esta política

[135] Cfr. Sayre-McCord, G., art. cit., p. 193, en esp. nota.

podría chocar con los intereses particulares, estaría condicionada por la escasez de información, etc. En definitiva, es imposible determinar si las condiciones realistas (en las que, no olvidemos, Gauthier pretende que su argumento sea válido) darían lugar a una triunfo o un fracaso de la MR. Como escribe Sayre-McCord:

> "En cualquier caso, cuando hablamos de personas reales en situaciones reales, está claro que el coste de defraudar puede estar, para algunos individuos, más que compensado por los beneficios esperados. Quienes sean capaces de ocultar con éxito su carácter (como lo son muchos) saldrán a menudo mejor parados desde el punto de vista auto-interesado si se vuelven (o siguen siendo) inmorales. Quizá en una comunidad compuesta por personas perfectamente racionales el riesgo de ser inmoral fuese más substancial. Pero incluso para el agente perfectamente racional que vive entre sus iguales, la inmoralidad será racional si defraudar es una opción viable. Así pues, no se puede condenar en general la inmoralidad acusándola de irracionalidad."[136]

Poco ayuda la translucidez, por tanto, a clarificar el problema de la racionalidad del cumplimiento de los contratos. Tanto desde el punto de vista intuitivo (representado por Sayre-McCord), como desde un enfoque técnico (intentado por Franssen[137]), la cuestión queda abierta. No se puede afirmar

---

[136] Sayre-McCord, G., art. cit., p. 194.

[137] Ahorramos al lector la enojosa (por compleja y detallada) argumentación de Franssen. Diremos que su conclusión se solapa con la de Sayre-McCord, como muestra el siguiente texto: "Para concluir: es dudoso si la translucidez podría servir para hacer que dos actores enfrentados a un Dilema del Prisionero logren una mutua cooperación, esto es, conformidad por parte de ambos con los términos de los contratos pactados. Los dos actores se enfrentan a un problema de teoría de juegos, pero la teoría clásica de Nash de juegos bi-personales no-cooperativos de suma no-cero no contiene una respuesta a su problema. Las teorías recientes sobre la deliberación dinámica sí permiten a los actores derivar la cooperación mutua como solución racional, pero las condiciones bajo las cuales se obtiene este resultado —referidas al modo de razonar de los actores y a los aspectos de la situación que deben ser públicamente conocidos (que incluyen el valor exacto de las "leyes psicológicas")— son exigentes. E incluso si se aceptaran los supuestos generales de estas teorías, habría aún que ajustarse a otros requisitos específicos de cada una de ellas."

con certeza que *cualquier* agente racional tenga una razón de auto-interés para convertirse en un agente moral. Pero tampoco se puede afirmar que *nadie* tenga esa razón. Sayre-McCord habla de tres clases de personas que encontrarían racional elegir *ser* morales (y no meramente aparentarlo): quienes fueran permanentemente translúcidos y se encontrasen en una comunidad de agentes morales; aquellos para quienes defraudar tiene un significativo coste personal; y quienes tengan la capacidad de participar en una comunidad moral y abrazar ciertos fines valiosos ajenos al auto-interés.

Si somos optimistas respecto a la magnitud de estos grupos de personas (especialmente respecto al tercero), puede aceptarse que la MR acabaría triunfando. En otro caso —y las premisas de la teoría con obligan a limitar ese optimismo— la conclusión será negativa: sólo en condiciones extraordinariamente contrafácticas parece racional adoptar una disposición sincera como la MR[138].

No obstante esta conclusión parcial, vamos a aceptar provisionalmente que la hipótesis de la translucidez validase el argumento segundo de Gauthier y que se demostrase la racionalidad de restringir el afán maximizador cuando existe la posibilidad de la cooperación[139]. Esta concesión momentánea nos permite proseguir con la discusión del razonamiento de Gauthier.

(Franssen, M., art. cit., pp. 267-268).

[138] Lo cual es admitido con toda naturalidad por Gauthier. Cfr. *MA*, pp. 181-182.

[139] En cualquier caso, la inoperancia de la hipótesis de la translucidez no invalida —creemos— los argumentos de Gauthier expuestos al final del sub-epígrafe anterior. Que no se pueda *demostrar* la racionalidad de adoptar la disposición MR, no significa que esa disposición no sea la que efectivamente resulta maximizadora de la utilidad (al igual que la imposibilidad de motivar racionalmente a un Ulises desatado para que pase de largo ante la isla de las sirenas no significa que eso no sea lo mejor para él). Por tanto, la crítica al expediente de la translucidez no afecta, en principio, a la tesis de fondo sobre la racionalidad de cumplir los acuerdos pactados. En este mismo sentido se expresa Gauthier cuando escribe: "Aunque la rentabilidad real de la cooperación condicional depende tanto de las habilidades como de la proporción de cooperadores entre la población, sin embargo, la rentabilidad potencial de la disposición no tiene un fundamento empírico, sino que refleja la estructura lógica de la interacción. Idealmente, un individuo cuyo objetivo es egoísta —obtener tanto como sea posible para sí mismo— debe esperar mejor resultado no como egoísta, sino como cooperador." ("The Incompleat Egoist", en Gauthier, D., *Moral Dealing*, cit., pp. 234-273; p. 265).

e) Qué clase de disposición es racional adoptar.-

Es una constante en el contractualismo y aún en el convencionalismo, definir la justicia por referencia a los pactos: en ausencia de cualquier otro criterio moral, lo justo se define simplemente como el cumplimiento de los acuerdos concertados voluntariamente. Sin embargo, los argumentos ofrecidos hasta aquí tratan el cumplimiento con independencia de la justicia. Gauthier intenta mostrar que es racional adoptar la disposición MR, es decir, disponerse condicionalmente a basar las propias acciones en una estrategia conjunta *sin* considerar si una estrategia individual produciría mayor utilidad esperada. Pero, como la adopción de esa disposición está a su vez basada en una decisión paramétrica maximizadora de la utilidad, no se ha apelado en ningún momento a la justicia. Hasta ahora, no se ha aclarado por completo (pese al lenguaje que en algún momento hemos utilizado, siguiendo a determinados críticos) la precisa conexión entre racionalidad y moralidad.

Para aclararla hay que concretar las características de la disposición condicional MR. Es importante advertir que para llegar a esa concreción nos elevamos de nuevo a un plano de racionalidad ideal del que imperceptiblemente habíamos descendido para discutir la hipótesis de la translucidez. La teoría del cumplimiento opera, así, en dos niveles o momentos: un primer momento centrado en la relación acuerdo-cumplimiento, en el que hay que hacer ver que los acuerdos pactados en una situación ideal con estructura del Dilema del Prisionero —como es el estado de naturaleza— se cumplirían después en las ocasiones en que su aplicación es relevante, es decir, aquellas situaciones en que la cooperación es posible, pero no es un resultado en equilibrio, con lo que su consecución parece exigir de las partes un comportamiento imposible de inscribir en una teoría de la racionalidad como maximización. En este primer nivel el acento se pone en mostrar que es racional para un individuo real[140] cumplir lo pactado (incluso aunque se trate de un acuerdo en condiciones ideales, siempre que el agente pueda identificarse con el resultado del mismo).

[140] "Real" —o, mejor, aproximadamente realista— por lo que se refiere a algunas características psicológicas y disposiciones racionales relevantes para el problema de la conformidad.

Por esa razón, los argumentos tienden a enfatizar las condiciones y dificultades reales del cumplimiento de los contratos cooperativos en general.

El *segundo momento* se centra en la *relación cumplimiento-justicia* y su objetivo es establecer normativamente los límites estrictos dentro de los cuales es racional (idealmente racional) cumplir un acuerdo. Lo que interesa ahora no es tanto mostrar el nexo entre auto-interés racional y cumplimiento, cuanto recobrar la noción de acuerdo imparcial y optimizador para poner el cumplimiento en función de ella. Si en el primer momento se pone en relación el cumplimiento con lo pactado (sin cuestionar explícitamente el contenido del pacto); en este segundo momento se pone en relación con la negociación misma, pues sólo una estrategia conjunta acordada tras una negociación racional (regida por el principio *minimax*) será cumplida por agentes perfectamente racionales. A la vez, las partes de la negociación sólo concertarán un acuerdo que pueda ser cumplido *ex post* por agentes perfectamente racionales, esto es, un acuerdo basado en el principio *minimax* o, en otras palabras, un acuerdo justo, en el sentido de imparcial (*fair*). De este modo, la teoría del cumplimiento racional y la teoría de la negociación se retro-alimentan, y forman un todo que no puede (a pesar de nuestro esfuerzo analítico) entenderse por separado.

Esta observación debe hacer ver la relevancia de la teoría del cumplimiento para la teoría de la negociación, pero, al incidir también en la influencia inversa, nos ayuda a fortalecer los argumentos en favor de la cooperación. Porque gran parte de las críticas expuestas contra la MR suponen un cálculo relativamente cicatero de los beneficios de la cooperación y un cálculo relativamente grande de los beneficios del fraude. Pero si la disposición a cumplir pactos se circunscribe únicamente a aquellos acuerdos perfectamente imparciales y óptimos, entonces, se incrementa considerablemente el beneficio de la cooperación y aquellos argumentos críticos pierden parte de su fundamento.

Además, cuanto más grande es el beneficio esperado de la cooperación, más débiles —y fáciles de alcanzar— son los requisitos necesarios para

convertirse en un MR[141]. Por esta razón, Gauthier piensa que su argumento en favor de la MR defiende, implícitamente, una concepción determinada de la misma: una concepción que denomina "cumplimiento estricto" (*narrow compliance*) y que se definiría como la disposición condicional a cooperar *únicamente* en aquellas prácticas y actividades que produzcan resultados óptimos e imparciales (o que se aproximen mucho a ellos)[142].

En apoyo del cumplimiento estricto, Gauthier argumenta que esa disposición puede afectar las oportunidades del agente que la adopta, de modo que domine a una disposición menos estricta. Según Gauthier, los agentes dispuestos a cooperar en empresas que no les ofrezcan un resultado óptimo, son blanco de la explotación, pues otros agentes les propondrían cooperar "a cambio" de una utilidad sólo algo superior al pago de la no-cooperación, aprovechando para sí la mayor parte del beneficio cooperativo. Quien abraza la disposición estricta siempre (y solamente) aceptará estructuras cooperativas basadas en el principio de concesión relativa *minimax*, que ningún agente perfectamente racional puede recusar. La cooperación se producirá así sobre la base de una racionalidad igual y será imparcial con todos. Gauthier añade que al rechazar otros términos para la cooperación, el agente estricto no disminuye sus perspectivas de participación con otras personas racionales —ya que suponemos que *todos* son perfectamente racionales y reconocen que esto es lo máximo que pueden exigir de un socio— mientras asegura que quienes traten de llegar a acuerdos imparciales no tendrán la oportunidad de cooperar, con lo

[141] Si los beneficios de la cooperación son escasos en relación a la no-cooperación, convertirse en un MR sólo será racional si se supone un alto grado de transparencia en los agentes. Pero conforme ese beneficio crece, el grado de transparencia (o translucidez) necesario será menor, con lo que la contingente decisión de convertirse en un MR gana plausibilidad. Sobre esto, cfr. Gauthier, *MA*, p. 178.

[142] Traduciremos *narrow compliance* indiferentemente como "cumplimiento estricto", como "disposición estricta (al cumplimiento)", o como "pre-disposición estricta". Creemos que el sentido del término (así como el de su antagonista, la pre-disposición "amplia" al cumplimiento) queda claro en las explicaciones del texto. Este sentido hace una cierta violencia al significado usual de estas palabras en castellano. El sistema que hemos elegido de usar varios vocablos, según el contexto, para traducir las expresiones técnicas de Gauthier tiende a atenuar esa violencia. Esperamos que no suponga una excesiva pérdida de precisión.

que su disposición será costosa y, por tanto, irracional[143].

Kraus y Coleman —y, entre nosotros, J.C. Bayón Mohíno— percibieron la importancia que tiene para el argumento de Gauthier la plausibilidad del cumplimiento estricto. Si éste es sustituida por una disposición amplia (aceptación y cumplimiento de acuerdos beneficiosos, aunque no sean exactamente imparciales y óptimos), entonces el contractualismo de Gauthier carece de todo valor normativo, ya que, como él mismo sugiere, incluso nuestras "malas" sociedades representan una cierta ventaja respecto al estado de naturaleza. Si la racionalidad no nos exigiera un compromiso con la imparcialidad, entonces *todas* nuestras sociedades estarían racionalmente justificadas y *ningún* cambio en ellas lo estaría, en sentido estricto.

Por eso, los comentarios sobre este punto merecen especial atención. Nos centraremos en particular en los de Kraus-Coleman y Peter Danielson (Bayón Mohíno refleja, más o menos, las tesis de estos autores).

Kraus y Coleman plantean un razonamiento sencillo que critica eficazmente la racionalidad del cumplimiento estricto, pero a base de introducir una premisa ilegítima. Sostienen que la igual racionalidad de las partes (a la que Gauthier apela) sería relativamente irrelevante a la hora de determinar la racionalidad de la disposición estricta o amplia. Lo relevante sería la "ventaja" de que cada quien gozase en la negociación y, eventualmente, la distribución previa de agentes estricta o ampliamente cumplidores[144]. Obviamente, la premisa ilegítima es la idea de "ventaja". Kraus Y Coleman hacen girar su argumento sobre un concepto desterrado por la propia lógica del contrato social. Si se refieren a la "dotación natural" de los agentes (de la que hablaremos en el punto siguiente), ésta se tiene en cuenta, desde luego, en la negociación; pero, ya vimos anteriormente que la diferencia de dotación natural no implica imparcialidad en el resultado de la negociación. Por tanto, entre

[143] Cfr. Gauthier, D., *MA*, pp. 178-179.

[144] Cfr. Kraus, J.S. y Coleman, J.L., "Morality and the Theory of Rational Choice", *Ethics*, 87 (julio, 1987), pp. 715-749; p. 744. Reimpreso en P. Vallentyne (ed.), *op. cit.*, pp. 254-290.

agentes igualmente racionales, no es posible hallar ninguna fuente de esa "ventaja"[145].

Contando con esa distinción entre las personas, Kraus y Coleman concluyen que una disposición estricta no sería racional ni para los aventajados (pues con ella perderían la posibilidad de aprovecharse de los acuerdos leoninos a su favor), ni para los desaventajados (pues les obligaría a renunciar a muchos acuerdos ventajosos, aunque injustos). Sólo en el caso de que la decisión de uno de los desaventajados sea decisiva —en el sentido de que completa el número de cumplidores estrictos que hace que sea excesivamente costoso para los aventajados intentar concertar acuerdos imparciales— habrá una razón para elegirla. Pero eso requeriría una población dada de cumplidores estrictos. Jamás se podría explicar por qué el primer desaventajado se tornó uno de ellos.

Bayón Mohíno reproduce el argumento de Kraus y Coleman[146], e infiere —creemos que con acierto— que la verdad de la tesis que defiende la MR (especialmente en su versión de cumplimiento estricto) es "puramente contingente"[147] y, por otro lado, conviene con Kraus y Coleman que al desarrollar el cálculo sobre la racional de adoptar la disposición que Gauthier defiende, se entraría en un círculo vicioso: es racional convertirse en un MR estricto sólo si ya hay un número determinado de agentes así dispuestos, pero nunca se alcanzará tal número si este cálculo es reproducido por todos los individuos.

Como alternativa a la versión de la cooperación condicional defendida por Gauthier, Peter Danielson ha propuesto la "cooperación recíproca"[148]. La

---

[145] ¿Habremos de recordar una vez más el inicio del capítulo XIII del *Leviathan*, que establece una premisa casi insustituible del contractualismo, "La Naturaleza ha hecho a los hombres tan iguales en las facultades físicas y mentales ..."?

[146] En parte con un objetivo diferente, el de criticar la salvaguardia lockeana y su papel en la teoría de Gauthier (Cfr. Bayón Mohíno, *op. cit*, pp. 178-179). Su crítica a la maximización restringida se encuentra más abajo, pp. 182 y ss.

[147] Cfr. Bayón Mohíno, *op. cit.*, p. 183.

[148] Cfr. su "The Visible Hand of Morality", *Canadian Journal of Philosophy*, vol. 18, nº 2 (junio 1988), pp. 357-384, sección III.

cooperación recíproca es definida como una meta-estrategia que domina a la cooperación condicional. La propuesta de Danielson tiene especial interés porque acepta, en general, el argumento de Gauthier. Admite que es racional restringir la maximización mediante la adopción de una disposición no-maximizadora. Pero, una vez decidido esto, cabe tratar las distintas opciones disponibles (las distintas versiones de la MR) como meta-estrategias entre las que se puede elegir conforme a la teoría de los meta-juegos[149]. Pues bien, la meta-estrategia "cooperación condicional" es estrictamente dominada por la meta-estrategia "cooperación recíproca", que es idéntica a la cooperación condicional excepto en que permite defraudar a los cooperadores *incondicionales* (en el caso de que existan). Desde un punto de vista estrictamente maximizador, la cooperación recíproca de Danielson se impondría a la cooperación condicional de Gauthier[150].

El argumento de Danielson no es definitivo por diversas razones[151], pero el propio Gauthier lo acepta como un intento de mejorar su teoría y, en cierta manera, una defensa indirecta del cumplimiento estricto[152]. De hecho, el cooperador recíproco es un cooperador más estricto incluso que los individuos gauthierianos, aunque no en el mismo sentido que éstos: al negarse a cooperar con los cooperadores incondicionales, el cooperador recíproco no sólo castiga a los MD (con su no-cooperación), sino también a quienes cooperan con ellos. De este modo, si las disposiciones del agente influyen en las disposiciones de los demás, o son interdependientes —lo cual es un

---

[149] Como obra clásica sobre meta-juegos puede citarse Howard, N., *Paradoxes of Rationality. Theory of Metagames and Political Behavior*, Cambridge (Mass.), MIT Press, 1971.

[150] Sobre las diferencias entre ambas meta-estrategias, y las razones para preferir la cooperación recíproca, ver Danielson P., art. cit., p. 380.

[151] Entre otras, porque requiere la existencia de varias meta-estrategias MR alternativas, incluida la disposición incondicional a cooperar, que marca la diferencia entre los cooperadores condicionales y los recíprocos; si esta disposición incondicional no existiera, no habría diferencia entre la propuesta de Danielson y la de Gauthier. Por otro lado, el propio Danielson no se atreve a afirmar que en las condiciones más realistas de la translucidez la cooperación recíproca venciese a la condicional, como lo hace en el aséptico meta-juego entre meta-estrategias; etc.

[152] Cfr. "Moral Artifice", *Canadian Journal of Philosophy*, vol. 18, nº 2 (junio 1988), pp. 385-418; pp. 400-101.

presupuesto de los argumentos que manejamos en esta sección— la MD (así como la cooperación incondicional) tendrán cada vez menos justificación[153].

No obstante, Gauthier recela de la propuesta de Danielson, a causa de su excesiva radicalización de la demanda de racionalidad: los cooperadores recíprocos no exhiben el menor sentimiento o el menor afecto. Su disposición a sacar partido de los cooperadores incondicionales es incontestable desde el punto de vista de la maximización —y, en ese sentido, plausible desde las premisas del contractualismo—, pero denota una carencia completa de preferencias afectivas hacia otros, o hacia un modelo determinado de sociedad. Este tipo de preferencias no está prohibido, en principio, por el requisito del auto-interés. De hecho, Gauthier cuenta con ellas para justificar una disposición hacia la cooperación algo más que simplemente recíproca.

El desarrollo y crítica de la idea de la maximización restringida difumina, como vemos, sus perfiles; aunque mantiene su fundamento esencial. Si no se ha podido ofrecer una justificación decidida de la maximización restringida como disposición estricta al cumplimiento, tampoco se ha podido refutar definitivamente su defensa.

Con todo, las dudas sembradas deberían hacernos reflexionar sobre la posibilidad de ofrecer una base más sólida para justificar el cumplimiento de los acuerdos pactados. Este punto queda tan abierto en *MA,* que da la sensación de que incluso Gauthier estaba pensando en su maduración posterior. Y ese desarrollo vendrá de la mano de un crítico que, en la línea de Danielson, acepta lo esencial del proyecto de Gauthier, pero procura fortalecerlo con una explicación más solvente de la racionalidad de cumplir los acuerdos. Se trata de Edward McClennen, cuya propuesta transciende el problema de qué

[153] Hay que advertir, siguiendo a Danielson (art. cit., p. 382), que este argumento encierra el peligro de un recurso infinito. Si es racional la cooperación recíproca porque "castiga" a quienes cooperan con los MD, tal vez sería aún más racional una cooperación recíproca estricta que castigase tanto a los que cooperan con los MD (los incondicionales) como a quienes cooperan con ellos (los cooperadores condicionales); y a su vez, otra aún más estricta que castigase a quienes cooperan con estos últimos (es decir ¿los cooperadores recíprocos!); y así sucesivamente hasta llegar a la cooperación reflexiva, en la que cada agente coopera con él mismo (y sólo con él), que sería la única situación estable.

disposición es racional adoptar para devolvernos a la cuestión sobre la naturaleza exacta de las "disposiciones racionales".

f) ¿Disposiciones, resoluciones o planes?.-

Lo primero que hemos de advertir es que McClennen no es el único crítico que pretende re-formular la idea de la maximización restringida manteniendo su fundamento pero desprendiéndose de la problemática noción de las "disposiciones racionales". Sí es, desde luego, quien más lejos y con mayor rigor ha llevado esta empresa y, por otro lado, uno de los teóricos que más ha influido en la evolución del pensamiento de Gauthier posterior a *MA* sobre este tema.

Entre quienes intentan un planteamiento diferente de la posibilidad de restringir el comportamiento maximizador deberíamos incluir a algunos críticos que ya hemos mencionado en relación con otros aspectos. Hemos visto que Danielson identifica las disposiciones con meta-estrategias (y así puede reconducir la discusión de Gauthier al campo algo más seguro de los meta-juegos). El mismo recurso emplea Barragán, aunque ese tratamiento no le lleva a ser tan condescendiente como Danielson con las tesis de Gauthier. Franssen no es explícito al respecto, pero también podría incluirse en el grupo de los "meta-estrategas", teniendo en cuenta su enfoque "ortodoxo". Por su lado, Varoufakis no hace una propuesta concreta, sino que sugiere distintos modos de "elección desviada", sólo justificables desde una concepción de la racionalidad superadora —o al menos distinta— de la concepción maximizadora.

Edward McClennen representa, en este conjunto de enfoques revisionistas, una aportación capital. Su crítica de la MR se inscribe en un proyecto global de análisis de la decisión individual a lo largo del tiempo, o "decisión

dinámica"[154]. Se trata, obviamente, de introducir las intuiciones de Gauthier —esencialmente correctas, según McClennen[155]— en el marco teórico de las que (a falta de un nombre mejor) podemos llamar en castellano "resoluciones" (*resolute choice/s*)[156].

La idea de una "resolución" es introducida por McClennen con el objetivo de superar las limitaciones aparentes de la decisión paramétrica en situaciones en las que hay que contar con el cambio de preferencias, preferencias futuras, realización de planes, etc. Estas limitaciones han quedado bien de manifiesto en los análisis de los teóricos que hemos denominado "ortodoxos". McClennen cree que la incapacidad de la decisión paramétrica para resolver coherentemente los problemas de elección dinámica se debe a tres características esenciales del razonamiento paramétrico: "(1) se presume que el agente tiene una ordenación de preferencias anterior, especificable sobre el conjunto de todos los posibles resultados de una acción, (2) la decisión racional consiste en seleccionar un curso de acción factible cuyo resultado asociado sea máximamente preferido y (3) el conjunto de consideraciones anteriores que condicionan cualquier momento de la decisión sirve solamente para restringir el conjunto de acciones factibles —no para modificar de ningún modo la

---

[154] Cuyo resultado se publicó en 1990: E.F. McClennen, *Rationality and Dynamic Choice: Foundational Explorations*, Nueva York, Cambridge U.P. Aunque algunas de las ideas esenciales ya se habían publicado en 1985, en "Prisoner's Dilemma and Resolute Choice", en Campbell y Sowden (eds.), *Paradoxes of Rationality and Cooperation*, Vancouver, University of British Columbia Press.

[155] McClennen sostiene, por ejemplo, que Gauthier logra mostrar claramente que los agentes auto-interesados *querrían* desarrollar una capacidad para interactuar cooperativamente entre ellos; pero no logra mostrar tan claramente como sería deseable que *puedan*. En un sentido parecido escribe en la p. 103 de su artículo (cit. en la nota siguiente): "Claramente será beneficioso para un agente —independientemente de qué disposición decida cultivar él mismo— animar a los demás para que desarrollen una disposición sinceramente cooperativa. Pero es menos obvio que sea racional cultivar esa disposición dentro de sí mismo", y, más abajo (p. 108): "Sigo convencido de que el tipo de teoría contractualista moral con la que Gauthier está comprometido es defendible. Lo que propongo es un enfoque alternativo (o reinterpretación de su argumento), que mantiene sus conclusiones sobre la racionalidad de actuar en las estructuras cooperativas."

[156] Cfr. McClennen, "Constrained Maximization and Resolute Choice", en E.F. Paul *et al.* (eds.), *The New Social Contract*, cit., pp. 95-118.

ordenación de preferencias del agente"[157]. La lógica del razonamiento paramétrico —representada por estos tres supuestos— hace difícil que las decisiones tomadas extiendan su fuerza a lo largo del tiempo. "En un contexto paramétrico, parece que la decisión de adoptar una disposición en un tiempo pasado t no tiene capacidad de mantenerse —no puede tener influencia sobre la cuestión de qué debe hacer el agente aquí y ahora, en el tiempo $t+1$"[158]. La elección paramétrica parece "indeformable" por las disposiciones.

Si se acepta la lógica del razonamiento paramétrico, el enfoque de Gauthier —basado en ella— no se sostendría. De ahí que McClennen, intentando mantener las conclusiones de *MA*, ofrezca su teoría de las resoluciones.

La lógica paramétrica se basa en la idea de que el agente ordena preferencialmente *los resultados* (y que su decisión está condicionada por esa ordenación). McClennen se pregunta simplemente si el agente no podría ordenar preferencialmente (y decidir entre) acciones, en vez de resultados. En este caso, no habría nada contradictorio en cooperar: el agente podría aducir que *prefiere* la acción cooperativa a la individual, aunque, en abstracto, pueda reconocer que el resultado de emplear una estrategia individual sería más beneficioso.

Lo que tiene que explicar McClennen es cómo es posible esa preferencia por una acción. Para ello considera que los agentes son seres que permanecen en el tiempo (en vez de considerarlos, como parece hacer Gauthier en *MA*, como seres "sucesivos", pero independientes unos de otros, que deciden paramétricamente en cada momento del tiempo). Estos seres deliberan, no tanto sobre los resultados de la acción inmediata, cuanto sobre el resultado final de distintos *planes* posibles, cada uno de los cuales se entiende como una secuencia o serie de elecciones. Una vez elegido un plan, las acciones que lo implementan son preferidas en sí mismas, con independencia del resultado que, en ese momento de la ejecución del plan, puedan tener. El mismo agente, enfrentado *ex novo* con esa elección (que ahora es parte de un plan) elegiría sin

[157] McClennen, E., art. cit., p. 105.

[158] McClennen, E., art. cit., p. 106.

duda la acción maximizadora de su utilidad; pero, al enmarcarse en un plan, elige la acción que lo implementará, según lo decidido anteriormente. Este tipo de decisiones son "sensibles al contexto"[159]. Y los agentes que las llevan a cabo reciben el nombre de "decisores resolutos" (o resueltos).

Es evidente que los agentes capaces de adoptar y llevar a cabo planes obtienen mejores resultados que los que deciden paramétricamente en cada momento. Y ese mejor resultado no se mide en relación a una decisión *ex ante*, como en el argumento de Gauthier en favor de la MR, sino en relación al interés que tienen todos y cada uno de los "yoes" sucesivos a lo largo del tiempo (representantes de la continuidad del agente)[160]. De esta forma, las incoherencias entre las decisiones paramétricas sucesivas, o entre una decisión paramétrica y las obligaciones de seguir una disposición, desaparecen.

Esta breve exposición no hace justicia, desde luego, a la densidad del argumento de McClennen. Pero servirá al menos para distinguir su aroma y comprender el alcance de su conclusión: "Si el argumento que he construido tiene éxito, la capacidad de cada agente para adecuar secuencias de elecciones a los planes adoptados, será suficiente para hacer la cooperación posible"[161].

La propuesta de McClennen es susceptible de varias críticas, en las que no nos detendremos[162]. Nos parece más importante hacer notar que la teoría

---

[159] La razón de este nombre está clara: mientras que para la Teoría de la Decisión Paramétrica siempre es posible predecir qué decisión es racional (la que maximiza la utilidad esperada), desde el punto de vista de McClennen eso no es posible, pues la elección racional será la maximizadora *si se trata de una decisión aislada, pero si forma parte de un plan, entonces dependerá del mismo.*

[160] Cfr. McClennen, E., art. cit., p. 113.

[161] McClennen, E., art. cit., p. 113.

[162] La primera sería la clásica objeción de que los planes racionales están sujetos a la eventual flaqueza de la voluntad del agente. A esto McClennen responde que, frente al enfoque clásico, que vería en la flaqueza de la voluntad el resultado de la operación de una motivación racional presente *(opuesta a la motivación racional anterior), para su enfoque, la flaqueza de la voluntad habría de* entenderse como un fracaso de la motivación racional misma. Otras objeciones girarían en torno a la posibilidad de que McClennen estuviese pre-suponiendo un cambio en las preferencias, que hiciera la cooperación un resultado paramétricamente racional (una vez diseñado el plan, el agente prefiere cumplirlo debido a que valora a sus semejantes o la cooperación misma). Una objeción

de las "resoluciones" es el modelo tras el que Gauthier desarrolla su última concepción de la racionalidad del cumplimiento de los acuerdos.

En efecto, como apuntábamos arriba, la ambigüedad en que queda la teoría del cumplimiento del capítulo VI de *MA* parecía exigir una reformulación del problema, y ésta ha llegado de la mano de una conceptualización de la racionalidad deliberativa inspirada en la teoría de McClennen. Aunque su tratamiento pormenorizado nos alejaría en demasía del objetivo de este trabajo, podemos remitir a un artículo concreto, "Assure and Threaten"[163], en el cual Gauthier compendia los avances posteriores a *MA* en este terreno. En ese texto, la reflexión gira en torno a nociones que podemos considerar "internas" (frente a las razones "sociales" e interdependientes en que se funda la defensa de la MR en *MA*), tales como la coherencia con los "yoes" anteriores[164], la idea de credibilidad y de una "intención sincera", o el concepto de un "plan de vida" racional. Gauthier ofrece un detalladísimo análisis de cómo cabría entender la noción de un "plan racional" y cuál sería su alcance. Por así decir, Gauthier intenta establecer el fundamento filosófico que haría racionalmente admisible la idea de una "resolución" en el sentido de McClennen. Esto implica ofrecer una nueva visión de la racionalidad deliberativa, que sustituiría enteramente a la visión clásica de la racionalidad como maximización. El alcance de sus conclusiones no es espectacular, pero —a nuestro juicio— le permiten restablecer plausiblemente el argumento en favor del cumplimiento de los contratos pactados (siempre que respondan a un criterio mínimo de racionalidad[165]), incluso aunque en el momento de su ejecución la utilidad esperada

---

más cuestionaría la legitimidad de fundar la racionalidad de una acción en algo que sucedió en el pasado (el diseño de un plan), en vez de adoptar un punto de vista prospectivo, naturalmente asociado a una teoría instrumental de la racionalidad. Los textos citados de McClennen revisan y responden —creemos que con solvencia— a estas y otras críticas.

[163] *Ethics*, 104 (julio 1994), pp. 690-721.

[164] A través del concepto de "compatibilidad intencional" con las acciones racionales anteriores (Cfr. Gauthier, D., art. cit., pp. 702-703.

[165] Cfr. art. cit., pp. 716 y ss. En breve, el criterio es que, en conjunto, el agente calcule que los objetivos generales de su vida se alcanzan mejor si en ella existe el plan formado por la promesa-y-su-cumplimiento, que si no hubiera habido promesa alguna.

de la acción que requieren (ordenada por la estrategia conjunta) sea inferior a la producida por una acción alternativa.

g) La racionalidad de cumplir los pactos.-

Para concluir este punto, tal vez debiéramos repetir las palabras de Jean Hampton, "el jurado aún está deliberando sobre si es racional para los individuos adoptar la maximización restringida"[166]. Lo que quizá ya no se cuestione es que, efectivamente, algún modo de restricción a la maximización directa sería globalmente beneficioso en términos de auto-interés. Las correcciones más solventes a la maximización restringida en su versión de *MA* (Danielson, McClennen) aceptan esa premisa mayor. Las críticas más radicales, la cuestionan o la olvidan o, simplemente, reconocen que les atrapa en unas redes de las que no pueden salir, conduciéndoles de un dilema a otro. Bajo nuestro punto de vista, hay dos razones por las que se llega a este callejón sin salida, que son a la vez, las dos raíces de las críticas a la maximización restringida. De alguna forma, ya se han mencionado en los epígrafes anteriores, pero las resumiremos aquí a modo de recapitulación final: se trata de la incomprensión del alcance real de la idea de "adoptar una disposición" y del hecho de pretender un análisis excesivamente centrado en el Dilema del Prisionero bi-personal.

La primera de estas razones consiste, por tanto, en reducir la elección entre disposiciones a una elección (sobre resultados) en el tiempo t_1 que hay que mantener en el tiempo t_2, cuando el orden de preferencia de los resultados contraría las preferencias que determinaron la elección en t_1. Este tipo de análisis es tan generalizado —lo encontramos incluso en McClennen— que habrá que concluir que Gauthier no logra explicar suficientemente en *MA* el sentido de "adoptar una disposición".

[166] Hampton, J, "Two Faces of Contractarian Thought", en Vallentyne, P., *Contractarianism and Rational Choice*, cit., pp. 31-55; p. 41.

Al adoptar una disposición a la maximización restringida el agente aplica, en efecto, la racionalidad paramétrica, pero lo hace *por última vez*, de modo que ese paradigma de decisión ya no es más aplicable. Hablar de un orden de preferencias en t_2 simplemente contradice la decisión tomada en t_1 (y ello con independencia de que la maximización restringida permita el comportamiento directamente maximizador en muchos casos). La opción por la maximización restringida debe entenderse, creemos, como una decisión sin vuelta atrás en sentido estricto: el agente racional puede elegir cambiar su paradigma de racionalidad, pero siempre lo hace conforme al paradigma que está ejercitando; el agente puede considerar que la MR es más beneficiosa desde el punto de vista auto-interesado —correspondiente a la racionalidad maximizadora que le hemos adscrito "por defecto"— pero, una vez adoptada, sólo cambiará (otra vez) de paradigma si ese cambio es beneficioso *de acuerdo con criterio establecido por la MR* (el paradigma que ejercita en este momento), y la maximización directa *no es* beneficiosa desde el punto de vista de un MR. Los críticos están suponiendo, por tanto, que la opción por un paradigma de racionalidad es una especie de menú permanente, y que el agente puede, ante cada interacción, no sólo elegir una estrategia, sino elegir un paradigma de racionalidad conforme a criterios *siempre* maximizadores (lo cual es una suposición ilegítima). No parecen tomar en serio la idea de que en la interacción se eligen estrategias (conjuntas o individuales), pero no modelos de racionalidad. La decisión sobre el modelo de racionalidad es anterior lógicamente, y ya no influye en las decisiones particulares.

Una vez que el agente decide cooperar, el criterio de racionalidad es la optimización y la imparcialidad (se podría decir, si se quiere, que para él, ahora maximizar *significa* cooperar; pero esto da lugar a otros malentendidos). Ello no implica que, en el marco de la cooperación, no se acuda a la competencia cuando es necesario (igual que los maximizadores directos recurren a la coordinación y a la cooperación coactiva en su afán de lograr la maximización a largo plazo)[167]. Pero los objetivos y los criterios de racionali-

[167] Un ejemplo de esto —cuyas derivaciones nos llevarían, por cierto, a otros temas— es el mercado: desde el punto de vista de la cooperación (y nos atrevemos a decir que desde el punto de vista del contractualismo moral liberal en general) el mercado es un "juego" donde se permite

dad subyacentes han cambiado por completo. Lo que ocurre es que ese cambio se explica por referencia a las premisas minimalistas del contractualismo, y esa explicación (tal vez equívoca) ha dado lugar a la interpretación de la crítica.

Por eso la sugerencia de "Assure and Threaten" tiene gran importancia. Allí, Gauthier hace ver que la posibilidad de adoptar una disposición —con independencia de que la disposición adoptada lo sea porque su puesta en práctica cumplirá mejor el objetivo de "maximizar la utilidad de cada uno" (único criterio racional *ex ante*)— tiene su raíz en el hecho de que nuestra racionalidad deliberativa es "flexible"; que no somos "máquinas de maximizar a corto plazo", como parecen suponer algunos críticos. En definitiva, Gauthier está reclamando que, al optar por la MR (volviendo al lenguaje de *MA*), los individuos expresan una naturaleza distinta a la que les hemos supuesto como punto de partida:

> "Al hacer esta elección, [las personas] estarían expresando su naturaleza no sólo como seres racionales, sino también como seres morales. Si la disposición a decidir de modo directamente maximizador estuviera indeleblemente impresa en nosotros, no podríamos restringir nuestras acciones del modo que la moralidad requiere."[168]

La restricción del comportamiento maximizador es ella misma maximizadora; pero la *capacidad* de llevar a cabo esa restricción (que no tiene nada que ver con el *hecho* de llevarla a cabo, ni con el cálculo auto-interesado que determina la racionalidad de hacerlo) expresa nuestra naturaleza como agentes morales.

---

la libre interacción competitiva (maximizadora) *porque* se estima que es la institución más eficaz para la consecución de logros optimizadores e imparciales en ciertos ámbitos (típicamente, la producción y distribución de bienes estrictamente privados). Al igual que, para algunos, la moral es una "astucia de la razón maximizadora" para alcanzar mejor sus objetivos; para otros, el mercado puede verse como una "astucia de la razón cooperativa (o moral)" para alcanzar mejor sus objetivos. Nótese que no es un simple juego de palabras: en el primer caso, el objetivo es la maximización *individual*, en el segundo es la optimización *colectiva* imparcial.

[168] Gauthier, D., *MA*, p. 184.

Pero, como anunciamos, la incomprensión del sentido de "adoptar una disposición" no es la única raíz de las críticas. Aludíamos en segundo lugar a que el análisis tiende a ver la MR como una solución a un juego concreto: el Dilema del Prisionero único (no reiterado) entre dos personas. Ciertamente cabe esa interpretación, puesto que la estructura de interacción del estado de naturaleza se ha identificado con un Dilema del Prisionero, pero, en ese caso, la solución al mismo sería la teoría contractual como un todo, no el mecanismo concreto de la maximización restringida. En este momento del argumento contractualista, el Dilema del Prisionero sirve únicamente como punto de referencia para contrastar el resultado (en términos de utilidad) de adoptar la disposición MR.

Tal vez tenga razón Binmore cuando sostiene que Gauthier no habla de un verdadero Dilema del Prisionero en esta fase del argumento (ya que, desde el momento en que el resultado cooperativo es más preferido, deja de existir tal dilema, al menos en su formulación clásica). Pero esto es justamente lo que Gauthier pretende: disolver el dilema mediante la instauración de un modo de deliberación que, ante la *misma* matriz de pagos, logra un resultado más coherente con la maximización que la propia actitud directamente maximizadora. Evidentemente, entre cooperadores racionales ya no hay Dilemas del Prisionero en sentido estricto[169], porque el tipo de interacción (maximizadora no-cooperativa) que da lugar a los mismos ha sido sustituida por la cooperación.

La dialéctica maximización directa/maximización restringida puede tomarse, en definitiva, como un modo de explicar que nuestra racionalidad deliberativa nos aconseja y nos permite cumplir los acuerdos pactados si éstos se ajustan a ciertos criterios de racionalidad relacionados con nuestro autointerés.

A lo largo de este epígrafe hemos repetido varias veces que será racional cumplir los acuerdos optimizadores e imparciales (aquí, "optimizador" e

[169] Es decir, las situaciones con estructura Dilema del Prisionero, no representan ya tales dilemas.

"imparcial" representan esos criterios de racionalidad[170]). Hemos dado por supuesto que el resultado de una negociación racional (conforme con el principio *minimax*) lo es. Sin embargo, este supuesto está justificado sólo en parte, porque al estudiar la negociación racional dejamos entre paréntesis provisionalmente la determinación de la posición original de la negociación (aceptamos como "dada" la dotación de cada negociador); y esa posición tiene un efecto causal sobre el resultado de la negociación, de modo que no es indiferente que sea una u otra.

El último eslabón de esta cadena circular (que es, por tanto, el primero) será la determinación de la posición inicial de negociación capaz de hacer racional (óptimo e imparcial) el resultado de la negociación y, por tanto, su cumplimiento.

[170] Hay que notar que se trata de criterios de racionalidad colectiva (aunque afectan a la racionalidad individual de cumplir los acuerdos). La versión estrictamente individual de estos criterios puede verse en "Assure and Threaten", cit.

5. Contrato social y derechos individuales

El objeto de este punto es discutir las características de la posición inicial de negociación según una teoría moral contractual[171]. Al hablar de la "posición inicial" volvemos al estado de naturaleza, para determinar su contenido material —el enfoque del punto *2.* puede considerarse "formal", pues allí nos ocupamos exclusivamente de la *estructura* de la interacción natural. Se trata ahora de precisar la dotación natural de cada agente: lo que cada uno *lleva* a la mesa de negociación.

Lo que cada negociador aporta a la empresa cooperativa ha de haber sido adquirido y poseído por él; él ejerce sobre ello cierto poder exclusivo, que permite identificarlo como *su propiedad.*

En la medida en que tal propiedad es reconocida como base para la negociación por los demás futuros-cooperadores racionales (pues si no lo fuera, simplemente no negociarían con ese agente[172]), la misma puede considerarse legítima, y el poseedor puede reclamar su *derecho* (tal vez exclusivo) sobre ella.

Así, la configuración de la posición inicial de la negociación —que, por lo demás, no es sino una parte de la teoría de la negociación racional— nos pone en contacto con un grupo de teorías sobre la justicia política (unas contractuales, otras no) caracterizadas por apoyarse en ciertas tesis sobre los derechos individuales naturales. Este contacto con las teorías de los derechos naturales causa constantes malentendidos que debemos disipar desde el

---

[171] Hay que precisar que, aunque el contractualismo liberal emplea la fórmula de una negociación racional para explicar el proceso que da lugar a los principios de la justicia, el debate de la posición inicial tiene poco que ver con el problema de la determinación del "punto de no acuerdo" en la teoría de la negociación racional. Sobre este tema, puede verse Brian Barry, *Theories of Justice*, cit., pp. 31 y ss.

[172] Cfr., a este respecto, el argumento y ejemplos de J. Buchanan en *The Limits of Liberty*, Chicago, University of Chicago Press, 1975, pp. 17-18.

principio, de modo que se aclare nuestra comprensión del contractualismo moral liberal.

La teoría de la posición inicial *no es* —desde nuestro punto de vista— una teoría de los derechos naturales en el sentido normalmente asociado a pensadores como Locke o Nozick. Más bien se trata de un componente (fundamental) de una teoría de la negociación racional, entendida a su vez como parte esencial de una teoría del contrato moral. Para el contractualismo, *todos* los derechos son post-contractuales[173]. En el estado de naturaleza cada individuo "posee" su cuerpo y sus capacidades naturales —ello se deriva directamente del postulado individualista— pero eso no le otorga derecho exclusivo alguno. Todo derecho tiene su origen y fundamento en el contrato[174]. Esto no quiere decir que la estructura de derechos que el contrato engendra no informe, a nivel hipotético, las etapas ulteriores de la teoría (las que toman como premisa a los individuos poseedores) y, a nivel real, la sociedad justa. En este sentido (y sólo en este sentido) cabría hablar de una teoría de los derechos, concibiéndola siempre en el marco de (y subordinada a) una teoría contractual de la moral y la justicia[175].

---

[173] En este punto nos consideramos herederos directos de Hobbes: en el estado de naturaleza todos tienen derecho a todo, lo que equivale a decir que nadie puede reclamar derecho exclusivo alguno.

[174] En este punto, puede servir de ejemplo (aunque enseguida nos ocuparemos de distinguir teorías políticas de teorías morales) la distinción que hace Buchanan entre "distribución natural" y "asignación de derechos" (Cfr. *The Limits of Liberty*, cit., cap. 2; en esp. pp. 23-28). Y podría ofrecerse como ejemplo contrario la visión de Nozick, según la cual los derechos individuales están dados y el contrato originario (el nacimiento de las primeras asociaciones de protección) está encaminado a protegerlos o garantizarlos.

[175] En este punto seguimos a Gauthier (cfr. *MA*, p. 193) quien afirma que "podemos decir que la salvaguardia moraliza y racionaliza el estado de naturaleza, *pero sólo en la medida en que concebimos que el estado de naturaleza da lugar a la sociedad*" (énfasis mío). A su vez, Gauthier explica su punto de vista por referencia a la formulación hobbesiana de la primera ley natural: "Que todos los hombres deben esforzarse por conseguir la paz, *en tanto en cuanto tengan esperanza de obtenerla*" (énfasis mío).

Nótese que este enfoque causa que el contractualismo, frente a las teorías de los derechos, suela reclamar instituciones sociales fuertemente re-distributivas. Gauthier, por ejemplo, en su primera aproximación al problema de la herencia, opina que "la herencia como tal no existirá; es una práctica que personas mutuamente desinteresadas no tendrían razón para instituir ni aceptar" (*MA*, p. 300). Cfr., para un desarrollo de esta idea K. Sauvé, "Gauthier, Property Rights and Future Generations", *Canadian Journal of Philosophy*, vol. 25, nº 2, Junio 1995, pp. 163-176.

La estructura de derechos racionalmente determinada se impone sobre el estado de naturaleza "hobbesiano" y limita la posición inicial de negociación. Esta limitación aleja a la teoría de Gauthier de otros contractualistas contemporáneos —como, por ejemplo, Buchanan—, que no limitan en modo alguno la "distribución natural" que les sirve de posición inicial. Enseguida veremos que esta diferencia es, a la vez, efecto y causa de que el contenido del contrato de Gauthier sea no sólo político, sino además moral. Porque al restringir racionalmente la posición inicial, el resultado del contrato no sólo será mutuamente ventajoso, sino también imparcial o justo. Apelará al interés de cada persona, y a la vez establecerá una estructura social que nadie podrá recusar sobre la base de que viola alguno de sus derechos. Al excluir la posibilidad de transmitir los efectos de la predación o la coacción naturales a la sociedad, el contrato de Gauthier se convierte en un ideal normativo de justicia (un ideal moral), como jamás podría hacerlo un contrato constitucional del tipo defendido por Buchanan.

a) El papel de la salvaguardia.-

La teoría de la posición inicial ha de precisar qué estado de cosas *previo* serviría como punto de partida para una negociación racional y un acuerdo con

---

Por último, debemos advertir que nuestra coincidencia con Gauthier es sólo parcial: no compartimos su insistencia en que la salvaguardia (que impone ciertos límites, como veremos, a la estructura de derechos naturales admisible como base de la negociación) supone una "moralización" del estado de naturaleza. Con ese tipo de argumentos —apoyados muchas veces en afirmaciones ambiguas— se da pábulo a la confusión entre el contractualismo moral y una teoría moral basada en los derechos naturales. Por ejemplo, en la misma página 193 de *MA*, Gauthier escribe "Sin la expectativa de un acuerdo y de una sociedad, no habría moralidad, y la salvaguardia no tendría fundamento. Afortunadamente, la expectativa de una sociedad ya la damos por realizada; ahora nos interesa comprender el fundamento de la moralidad que la sustenta". Estamos de acuerdo con la primera frase; pero en la segunda parece sostener que la única fuente de moralidad en la sociedad basada en el contrato son los derechos naturales; y esto contradiría el argumento de una moral contractual. Sostendremos que la forma de eliminar esta ambigüedad es poner todo el énfasis en el primer tipo de afirmaciones, haciendo hincapié en que la moralidad y los derechos *nacen* del acuerdo hipotético. En este sentido, nuestra posición se encontrará más cerca de Buchanan que del propio Gauthier, quien precisamente critica la tesis de que sólo se puede hablar de derechos en la medida en que se ha realizado un acuerdo formal (cfr. *MA*, p. 199).

el que cualquiera pudiese estar conforme. La base lógica de esta teoría es que, si el estado de naturaleza es tal que la negociación y el pacto son rigurosamente imposibles, entonces nada de lo dicho hasta ahora tendría el menor interés. Pero, a la vez, como quiera que la cooperación implica una ventaja para cada uno en relación al estado de no-cooperación, hay que suponer que todo agente racional tiene interés en que el pacto se realice. Por lo tanto, todos tendrán interés en crear las condiciones que lo hacen posible, limitando, si fuera necesario, su libertad natural —bien entendido que ese límite, supeditado a la posibilidad del pacto, se respeta sólo en relación a aquellos con quienes se espera cooperar en el futuro.

Los límites que hay que imponer a la interacción natural para que configure una posición inicial de negociación aceptable se resumen en lo que Gauthier, siguiendo a Nozick, denomina "salvaguardia lockeana" [*lockean proviso*]. Esta salvaguardia garantizaría que en el estado de naturaleza "nadie debe empeorar la posición de otro salvo que ello sea necesario para evitar un empeoramiento de la propia posición"[176].

Antes de considerar el significado y contenido exacto de esta salvaguardia, quizá sea interesante observar se trata de una restricción de la interpretación que Nozick hace del límite en la adquisición de propiedad enunciado por Locke en el *Segundo tratado sobre el gobierno civil*. Locke se refiere a este límite principalmente en los parágrafos 27 y 33 del capítulo quinto del *Segundo tratado*. En el primero de estos párrafos Locke alude al límite en la apropiación de los productos del propio trabajo, diciendo que "... este trabajo, al ser indudablemente propiedad del trabajador, da como resultado el que ningún hombre, excepto él, tenga derecho a lo que ha sido añadido a la cosa en cuestión, *al menos cuando queden todavía suficientes bienes comunes para los demás*" (énfasis mío). El parágrafo 33 amplía esta misma idea a la apropiación de la tierra. En ese parágrafo 33 quedan claras dos cosas: primero, que Locke concibe su teoría de los derechos naturales para un mundo en el que la provisión de tierra (materias primas en general) se supone prácticamente

[176] Cfr. Gauthier, D., *MA*, p. 203.

infinita; segundo y más importante, que el fundamento de su tesis es que mediante la apropiación originaria y mejora de la tierra (o de cualquier otro bien muy abundante) no se perjudicó a los demás hombres, dado que "todavía quedaban muchas y buenas tierras, en cantidad mayor de la que los que aún no poseían terrenos podían usar". De esta idea subyacente extrae Nozick el argumento que le sirve para ofrecer su propia formulación de la salvaguardia: Nozick critica muy convincentemente la cláusula cautelar lockeana[177], y llega a la conclusión de que su único sentido coherente consistiría en garantizar que la situación de los otros no empeore. A su vez, esa garantía puede tener dos lecturas, una *fuerte* y otra *débil*, de las cuales sólo la débil evitaría la crítica de Nozick (que es una *reductio ad absurdum*):

> "Alguien puede empeorar por la apropiación de otro de dos maneras: primera, perdiendo la oportunidad de mejorar su situación con una apropiación particular o una apropiación cualquiera; y, segunda, por no ser ya capaz de usar libremente (sin apropiación) lo que antes podía. Un requisito *riguroso* de que otro no empeore por una apropiación excluiría la primera manera, si ninguna otra cosa compensa la disminución de oportunidades; así como la segunda. Un requisito más *débil* excluiría la segunda manera, pero no la primera [...]. Se puede sostener que nadie debe lamentarse legítimamente si se satisface la estipulación más débil."[178]

De hecho, sólo la salvaguardia débil permitiría la propiedad privada, mientras la versión fuerte la impediría completamente (siempre según la crítica de Nozick a Locke). Sin embargo, Gauthier cree que la salvaguardia "débil" defendida por Nozick es aún demasiado fuerte pues, en ocasiones, empeorar la situación de otros puede ser el único medio para evitar el empeoramiento propio y, en este caso, habría que relajar incluso la salvaguardia débil. De ahí

---

[177] Cfr. Nozick, R., *Anarquía, estado y utopía*, México, F.C.E., 1988, p. 177.

[178] Nozick, R., *op. cit.*, p. 177.

el concreto enunciado de la misma según Gauthier.

Hecho este pequeño "árbol genealógico" de la salvaguardia, debemos analizar ahora los límites concretos que impone a la interacción natural.

Gauthier descompone en dos partes las restricciones asociadas a la salvaguardia. La primera parte concierne a la coacción; la segunda al "empeoramiento" de otros propiamente dicho (derivado de la apropiación o de cualquier otra actividad). En cuanto a la coacción, debe estar desterrada del estado de naturaleza, puesto que nadie aceptaría un acuerdo cuya posición inicial fuese coactiva. Gauthier demuestra que la única manera de mantener el cumplimiento de tal acuerdo consistiría en extender la coacción a la sociedad. Su argumento emplea la noción de "transferencia improductiva" para explicar que la interacción coactiva daría lugar a cesiones de utilidad en la sociedad no requeridas por el principio *minimax*. Mediante la coacción, un agente puede detraer parte de los bienes producidos por otro en el estado de naturaleza. La posición inicial determinada por esa relación entre ambos influirá en el punto de demanda respectivo, lo cual influirá, a su vez, en el resultado de la negociación[179]. Este resultado exigirá que parte de lo que el agente coaccionado cedía al primero, se siga cediendo ahora en la sociedad, pero ¿qué justificaría esta transferencia, en ausencia de coacción? Se trataría de una transferencia de utilidad improductiva (gratuita, como una donación). El agente obligado a la misma no cumplirá con ella, porque sabe que el coste de la coacción necesaria para obligarle a hacerlo reinstauraría la situación natural, de

[179] La concesión se mide el relación al resultado no-cooperativo de cada negociador, pero si ese resultado es distorsionado por la violencia o la coacción, tal distorsión se reflejará en la estrategia cooperativa. El coaccionado tendrá razones para asentir a un acuerdo *formalmente* imparcial: derivadas del hecho de estar siendo coaccionado, pues el acuerdo, por leonino que sea, al menos elimina la coacción. Mas —por continuar con el argumento de Gauthier— una vez eliminada la amenaza, no será racional mantener los intercambios anteriores. Sabido esto, el coaccionador *no* pactará con el coaccionado. Es decir, una situación coactiva no servirá como base para un acuerdo racional. Ahora bien, la cooperación es beneficiosa *incluso* para quien tiene una posición ventajosa en el estado de naturaleza, luego —de acuerdo con nuestras premisas habituales— será racional desmantelar el sistema coactivo antes (entiéndase en sentido lógico) de iniciar la negociación.

la cual todos han considerado beneficioso salir[180]. Y, por la misma razón, si el agente anteriormente beneficiado de la coacción es completamente racional, admitirá este incumplimiento, porque reconocerá que la obligación que demandaba la transferencia improductiva estaba derivada de su coacción previa; la cual no habría sido aceptada como punto de partida de la negociación por ningún agente racional. De este modo, el resultado de la cooperación se irá moviendo hacia el resultado que *habría* producido una negociación basada en una interacción no coactiva.

Como conclusión del análisis de la coacción en el estado de naturaleza, Gauthier escribe:

> "Concluimos que la posición inicial de negociación, como punto de partida para la cooperación racional, no puede identificarse con la distribución natural, o resultado no-cooperativo. La distribución natural representa los efectos de la fuerza. Ni hemos mostrado hasta ahora, ni mostraremos, base alguna racional o moral para criticar los efectos de la fuerza considerados en sí mismos. Pero si consideramos la interacción natural en relación con la interacción de mercado o cooperativa, entonces la valoramos, no en sí misma, sino por su adecuación para determinar lo que cada persona trae al mercado o a la negociación que subyace a la cooperación. La valoramos como determinadora de la dotación inicial de cada persona. Y aquí hemos encontrado una base para criticarla y, de hecho, la rechazamos en la medida en que sea coactiva."[181]

La segunda parte de la salvaguardia se centra en la prohibición de "empeorar" la situación de otro excepto que sea necesario para no empeorar la

---

[180] El detalle de este argumento se puede ver en *MA*, pp. 194-197; su discusión y un aclarador gráfico, en *MA* pp. 227-231.

[181] *MA*, p. 198.

propia[182]. El significado de "empeorar" (o, para el caso, "mejorar") la situación de otro habrá de medirse por referencia a una situación-base en el estado de naturaleza. Gauthier sostiene que esa situación-base puede identificarse con el resultado que uno podría esperar en ausencia de los otros agentes relevantes[183]. En concreto, el comportamiento de cada agente quedará limitado por la salvaguardia si actúa de modo que el resultado (de la interacción) le proporcione a él mismo y a los demás una cantidad de utilidad mayor que la que esperarían en ausencia de interacción. Si esto no es posible, al menos intentará obtener para sí un resultado igual o mejor al esperado en ausencia de interacción, aunque para otros se derive una utilidad inferior a la que ellos esperarían en su ausencia.

El funcionamiento de la salvaguardia es teóricamente sencillo: aquellos que quieren ajustar su comportamiento natural a ella, simplemente limitan sus acciones a las permitidas. Esa limitación transforma completamente el estado de naturaleza hobbesiano, y permite la cooperación:

> "La salvaguardia desempeña para nosotros un papel más amplio y más básico [que para Locke]. La tratamos como una restricción general, mediante la cual podemos movernos desde un estado de naturaleza hobbesiano hasta la posición inicial para la interacción

---

[182] Es evidente que la coacción es una forma de empeorar la situación de otro no legitimada por la excepción de la "defensa propia". La hemos tratado independientemente sobre todo para seguir a Gauthier, aunque hay otras razones, porque su desgajamiento ayuda a entender el papel de la salvaguardia (ya que la prohibición de la coacción es un límite que tendemos a justificar más fácilmente) y permite relacionar el enfoque de Gauthier con el de Buchanan.

[183] Para ver algunos ilustrativos ejemplos, cfr. *MA*, p. 204, 207. Gauthier usa estos ejemplos para mostrar la base "natural" ("en modo alguno arbitraria", escribe en la p. 221) de un supuesto derecho exclusivo a las propias capacidades y talentos, así como al propio cuerpo (de donde luego deriva el derecho *sobre* los frutos de esas capacidades. Se justifica así la dotación básica de los individuos en el mercado (que hubo de ser supuesta en su momento) y la utilidad asociada al resultado no-cooperativo, supuesta en la teoría de la negociación. El fundamento para la identificación de las capacidades y talentos personales con una *dotación natural* no-arbitraria fuente de derechos exclusivos (no se trata de una identificación inmediata; estas capacidades podrían considerarse *comunes*), es que se trata de capacidades de las que uno puede siempre hacer uso (se encuentre solo o en una comunidad) y nadie puede usar en su ausencia (cfr. *MA*, p. 209).

social."[184]

Como añade Gauthier algo más abajo, "la salvaguardia convierte las ilimitadas libertades de la naturaleza hobbesiana en derechos exclusivos y obligaciones". La deducción de estos derechos es explicada muy largamente por Gauthier, quien parece sentirse obligado a ofrecer un fundamento moral para la afirmación inicial de Nozick en *Anarquía, estado y utopía*, "los individuos tienen derechos, y hay cosas que ninguna persona o grupo puede hacerles sin violar los derechos...". Sin embargo, el argumento es, para nosotros, simple; *y no incluye necesariamente una referencia a derechos pre-contractuales*, o que haya que suponer necesariamente como previos al pacto[185].

La premisa mayor del argumento es la igual racionalidad de las partes. Habría que añadir también una aproximada igualdad en las capacidades naturales (físicas y mentales), de modo que para todos es mucho más ventajoso cooperar que intentar maximizar su utilidad el un estado de interacción no-cooperativo. La primera premisa, ideal, dota al resultado de su componente normativo; la segunda, empírica, lo dota de su poder motivador. Dadas estas premisas, se pueden comenzar a acoplar las piezas del puzzle contractualista (ya

---

[184] Gauthier, D., *MA*, p. 208.

[185] Como ya hemos dicho, nuestra interpretación se basa en el texto de Gauthier y en el programa del contractualismo moral; pero nos sentimos en la necesidad de enfatizar este tipo de afirmaciones a causa de que, en su afán de aclarar la diferencia del contractualismo liberal con el utilitarismo y con contractualismo rawlsiano, Gauthier emplea un lenguaje tan cercano a la teoría de los derechos naturales que podría dar lugar a confusión. Nosotros tomamos como clave explicativa del argumento el siguiente párrafo de la p. 222 de *MA*, en diálogo con el teísmo asociado a las teoría clásicas de los derechos, como la de Locke: "El contractualismo ofrece una comprensión secular de los derechos. Pero la idea de una moral por acuerdo puede confundir, si se supone que los derechos son el resultado de un acuerdo. Si nuestra explicación fuera esa, deberíamos suponer los derechos determinados por el principio de concesión relativa *minimax*. Pero, como hemos visto, la aplicación de este principio o, en general, el nacimiento de la interacción cooperativa o de mercado, exige una definición inicial de los actores en términos de sus dotaciones iniciales, y hemos identificado los derechos individuales con esa dotación. Los derechos proporcionan el punto de partida del acuerdo, no su resultado. Son lo que cada persona aporta a la mesa de negociación, no lo que se lleva de ella". Estamos de acuerdo con el espíritu de este texto: el contractualismo ofrece una base secular; un fundamento exclusivamente racional de los derechos naturales *y* estos derechos no *equivalen* al resultado del acuerdo. Pero —como diría Buchanan— sí son *reconocidos* sólo tras el pacto inicial y, en este sentido, es difícil asentir a la literalidad de las frases de Gauthier en las que identifica la dotación inicial con los derechos individuales.

hemos anticipado esta construcción, y regresaremos a ella): sólo si se espera el cumplimiento, será racional negociar y pactar una estrategia conjunta; sólo si el pacto es imparcial y mutua e igualmente ventajoso se puede esperar el cumplimiento; sólo determinadas situaciones iniciales darán lugar a resultados imparciales y mutuamente ventajosos a través de un proceso de negociación racional; luego es racional aceptar los contratos que podrían ser resultado de una negociación racional a partir de una posición inicial adecuadamente construida, *y sólo esos*.

De este modo se justifica racionalmente, en el interés que cada agente tiene en su propio beneficio y, por ende, en la cooperación, la imposición de restricciones a la libre interacción natural; restricciones que se identifican con la salvaguardia, porque la salvaguardia lockeana evita el tipo de interacción natural que transmitiría un sesgo injustificado (parcial) al resultado del contrato.

Las restricciones que la salvaguardia impone pueden verse, si se quiere, como una estructura de derechos y obligaciones en el estado de naturaleza hobbesiano que permiten alcanzar una posición original desde la que la cooperación es posible. Pero se trata de una metáfora: en el estado de naturaleza no hay derechos propiamente dichos. Las restricciones que la salvaguardia impone son realmente —por decirlo así— un algoritmo corrector que los agentes racionales aplicarán a la posición inicial de negociación si realmente quieren producir una estructura de interacción cooperativa estable. Restricciones que, si bien no pueden considerarse *resultado* de la negociación o del pacto social (no son derechos convencionales), sí surgen del pacto, en la medida en que son un supuesto necesario del mismo, pero carecen de justificación fuera de él[186].

---

[186] Nuestra interpretación del papel de la salvaguardia puede asociarse a la esquematización del argumento gauthieriano que hacen Kraus y Coleman (en "Morality and the Theory of Rational Choice", cit., p. 721). Según Kraus y Coleman la articulación de los tres elementos esenciales del contrato de Gauthier se podría mostrar mediante un argumento en cinco pasos:

"1.- La negociación es racional sólo si la conformidad es racional.
2.- La conformidad es racional sólo si el resultado de la negociación es justo [*fair*].
3.- Luego, la negociación es racional sólo si es justa.
4.- Luego, las negociaciones racionales deben ser justas.

Nuestra interpretación del papel de la salvaguardia lockeana en el argumento contractualista, parece estar confirmada por autores como Braybrooke cuando, ante la posible objeción de que la salvaguardia introduce una restricción moral (ilegítima) en el estado de naturaleza, afirma que "en tanto afecta a la negociación, la salvaguardia simplemente impide que el proceso negociador quede sesgado por demandas mínimas inflacionadas para cubrir los recursos adquiridos mediante el expolio de otros. Es algo que las victimas pondrían como condición para aceptar un acuerdo voluntario"[187]. No obstante esta —creemos que correcta— comprensión del papel de la salvaguardia en la negociación, Braybrooke también denuncia que la tesis de que "la salvaguardia implica derechos", tiene la apariencia de una fuente de restricciones morales (que han de funcionar en la interacción cooperativa y de mercado) independientes de la negociación y el contrato[188]. Es justamente en previsión de esa fácil crítica que creemos que el énfasis ha de ponerse en que todos los derechos son post-contractuales, en vez de en la imagen de un estado de naturaleza "moralizado" por la salvaguardia.

La sospecha de que la salvaguardia puede ser una vía subrepticia para introducir restricciones morales ajenas al contrato constituye tal vez la crítica más común e inmediata. Pero no es la única. En los epígrafes siguientes intentaremos mostrar —como hemos hecho con las otras concepciones centrales de la teoría— la orientación de las más pregnantes, así como las re-interpretaciones de la cláusula lockeana que consideramos más acertadas.

5.- Luego, negociar desde cualesquiera ventajas injustas que se pudieran tener *ex ante* no es racional (la razón: negociar desde posiciones injustas arroja resultados injustos, que no es racional cumplir. Pero si no es racional cumplir, entonces no es racional negociar)."

[187] Braybrooke, D, "social Contract Theory's Fanciest Flight", *Ethics*, 97 (julio 1987), pp. 750-764; p. 755.

[188] Braybrooke no ha sido, lógicamente, el único teórico que ha percibido este problema. Como ejemplo de una visión aún más contradictoria con lo que creemos es el verdadero espíritu de la salvaguardia, puede verse Paul Torek, "Liberties, Not Rights: Gauthier and Nozick on Property", *Social Theory and Practice*, vol. 20, n º 3, otoño 1994, pp. 343-362. Torek llega a deducir "deberes morales" directamente de la salvaguardia.

b) Libertad y derechos en el estado de naturaleza.-

La configuración de la posición inicial de negociación —un elemento habitual en las teorías del contrato— desliza a Gauthier hacia el complejo problema del fundamento de los derechos naturales. Ya hemos explicado que, en nuestra opinión, tal deslizamiento es sólo aparente; pero ello no impide que la mayor parte de las críticas hacia la salvaguardia provengan de teóricos de los derechos, que leen esta parte del argumento contractualista casi con total independencia de sus otros componentes. Tal sesgo es, por sí solo, causa de la mayoría de las críticas, basadas en cierta incomprensión hacia el contractualismo.

Pero si el común origen de las críticas las hace relativamente homogéneas, en otro sentido son muy dispares, ya que se pueden encontrar entre ellas dos posiciones completamente opuestas: quienes consideran que la salvaguardia es demasiado "débil" y no es suficiente fundamento para los derechos naturales (Donald C. Hubin y Mark B. Lambeth, Paul Torek); y quienes, por el contrario, creen que la salvaguardia es un límite intolerable a la libertad natural —de modo que, literalmente, "viola" los derechos naturales— (Buchanan, Jan Narveson). Por lo demás, las interpretaciones y los intereses de los críticos son lo bastante variados como para que resulte difícil ofrecer un tratamiento conjunto. Nos limitaremos, por tanto, a completar la visión general que venimos presentando y a añadir, acaso, algunas pinceladas representativas de cada una de las posiciones, tal como las entendemos.

Los teóricos que critican la debilidad de la salvaguardia lockeana de Gauthier se basan en que la derivación de unos derechos de propiedad extensos y a la vez precisos —como los que Gauthier acaba defendiendo— exige un *fundamento más profundo que el ofrecido por la libertad en el uso de las* propias facultades y capacidades[189]. El propio Gauthier reconoce que los derechos naturales no pueden ser defendidos, en el sistema de Locke, sin apelar

[189] Paul Torek representa paradigmáticamente esta posición que podemos llamar "lockeana". Su tesis principal es que el estado de naturaleza de Gauthier o Nozick son excesivamente "libertarios". La deducción de la propiedad privada es falaz en estos casos: la salvaguardia por si sola no da lugar a derechos exclusivos de propiedad.

a su origen divino. Tal vez se encuentre ahí el ancestro remoto de este tipo de crítica. En concreto, Paul Torek es muy persuasivo en su de-construcción de la deducción "secular" de los derechos de propiedad a partir de la salvaguardia; pero, desafortunadamente, no ofrece una visión positiva de cómo habría de ser el estado de naturaleza, provisto de una "más rica fundamentación moral", para dar lugar a amplios derechos de propiedad exclusiva. Indudablemente, no se trataría ya de un verdadero estado de naturaleza —una noción heurística diseñada como premisa del contrato— sino de algo muy diferente. En este punto, el enfoque de Torek hace que su crítica sea irrelevante, a pesar de ofrecer argumentos muy plausibles sobre la dificultad de deducir derechos de propiedad permanentes mediante argumentos como los de Gauthier y Nozick[190]. Si la única fuente de los derechos hubiera de ser la salvaguardia por sí sola, es evidente que no se sostendrían, sino que imperaría el estado hobbesiano de completa libertad —lo cual es, por cierto, abiertamente reconocido por Gauthier. Los derechos se sostienen en función del pacto. Por si sirve una imagen: el pacto *petrifica* y hace estable una estructura de derechos que —suponiendo que llegara a nacer antes o con independencia del contrato, cosa de por sí imposible— duraría, en un puro estado de naturaleza, sólo un instante.

Donald C. Hubin y Mark B. Lambeth[191] representan una posición cercana a la de Torek en algunos aspectos, tales como el enfoque de su ensayo[192] y el diagnóstico sobre la debilidad de la salvaguardia. No obstante, en otros aspectos se sitúan en las antípodas de Torek porque, debido a un análisis tal vez excesivamente casuístico, afirman que la salvaguardia es

---

[190] Cfr. Torek, P., art. cit., p. 352.

[191] Hubin, D.C. y Lambeth, M.B., "Providing for Rights", en Vallentyne, P. (ed.), *COntractarianism and Rational Choice*, cit., pp. 112-126.

[192] Que queda patente en el siguiente aserto: "Piensen lo que piensen los teóricos de los derechos naturales de los aspectos contractualistas de la teoría de Gauthier, su defensa de los derechos representa una significativa contribución a la teoría de los derechos naturales" (p. 114).

demasiado fuerte o restrictiva respecto a ciertos derechos individuales[193]. Hubin y Lambeth objetan que la salvaguardia no sirve para evitar un número de interacciones que, sin "empeorar" la situación de nadie en términos de su utilidad personal, nos parecen intuitivamente contrarias a los derechos naturales[194]; en este sentido, sería necesaria una restricción mayor. También critican que la salvaguardia permita empeorar la situación de otro cuando, de no hacerlo el agente, lo hubiera hecho de todas formas otra persona. Se trata, como hemos dicho, de críticas puntuales basadas en un cuestionable juego casuístico que apela a las convicciones morales sociales más arraigadas[195]. Según nuestro criterio, los reproches de esta índole no afectan en absoluto al argumento contractualista.

Tal vez la salvaguardia pudiera ser modificada para adecuarse aún más a los límites que una negociación racional debe imponer a la situación inicial. Tal vez algunas sugerencias de autores como Hubin y Lambeth o Torek podrían ayudar en ese sentido; pero no creemos que, en lo fundamental, sus posibles innovacionaes no estén ya incluidas en una amplia comprensión de la idea de "no perjudicar a otro salvo que sea necesario para no sufrir perjuicio uno mismo".

En la dirección contraria, autores como Jan Narveson opinan, siguiendo la tesis de Buchanan, que la salvaguardia de Gauthier es innecesariamente restrictiva. El punto crucial de su argumento es que, en el camino (real o hipotético) del estado de naturaleza a la sociedad y, dentro de ella, en cada

---

[193] Cfr. Hubin y Lambeth, art.cit., pp. 116-117. Se trata, no obstante, de casos que podemos desechar, dado que, como los propios autores reconocen, los ejemplos que emplean para mostrar la "excesiva fuerza" de la salvaguardia sólo se darían en el marco de una determinada estructura social en la que pudiera hablarse de justicia, injurias, castigos o penas. Nada de esto cabe en el estado de naturaleza.

[194] El casuístico —y un tanto delirante— argumento puede verse en Hubin y Lambeth, art. cit., pp. 117-123.

[195] La mayoría de sus ejemplos conciernen a prácticas tales como hipnotizar a otros para sacar provecho sexual de ellos, educar programadamente a niños con el mismo propósito, etc. Aunque no dudamos del afán de rigurosidad del análisis de estos autores, hemos de reconocer que, al menos en este aspecto, sus objeciones apenas rozan (mucho menos cuestionan) el papel o el contenido de la salvaguardia en el argumento de Gauthier. Están criticando más bien —tal vez sin percatarse de ello— la concepción económica de la utilidad.

etapa histórica, los individuos siempre se ven a sí mismos inmersos en un progreso hacia mejor, porque, en sociedad "no sólo estamos mejor de lo que estaríamos de continuar en un estado de naturaleza, sino que, si todo va bien, cada uno de nosotros está mejor en cualquier momento dado de lo que lo estaba en el momento precedente. Aunque tu, con un punto de partida superior gracias a tu mayor esfuerzo predatorio en el estado anterior, vayas por delante de mí, yo siempre estoy por delante de donde habría estado en otro caso"[196]. Este argumento justificaría la aquiescencia de los perjudicados por la interacción natural. Incluso un contrato relativamente desventajoso para ellos es beneficioso en relación a lo que podrían esperar en otro caso. Por otro lado, se argumenta que, una vez alcanzado el pacto constitucional —que, por así decir, sacraliza la distribución natural (hobbesiana)— las sucesivas etapas de desarrollo post-constitucional (el intercambio de mercado, la interacción cooperativa, las instituciones sociales, etc.) tienden a minimizar los efectos de la predación natural. De este modo, cualquier restricción a la interacción natural está injustificada moral, política o económicamente.

Este tipo de argumento es atractivo porque toma en serio la idea de que los derechos son siempre fruto de un pacto, y que la interacción natural está completamente libre de restricciones. De alguna forma, Buchanan (y su abanderado, Narveson) podrían decir a Gauthier que ellos sí demuestran que un estado de naturaleza hobbesiano puede dar paso a la cooperación, mientras que Gauthier acaba por echar mano de un estado de naturaleza *cuasi*-lockeano, contradictorio con su proyecto moral[197].

Pero reconocer el mérito del argumento no significa admitir la validez de la crítica. Entre otras razones, porque el propósito de Buchanan es construir una teoría *política* y económica a partir de presupuestos hobbesianos; mientras que el de Gauthier es construir una teoría moral. Las necesidades teóricas son

---

[196] Narveson, J., "Gauthier on Distributive Justice and the Natural Baseline", en Vallentyne, P., *Contractarianism and Rational Choice*, cit., pp. 127-148; p. 143.

[197] En este sentido, Narveson escribe que "el punto de vista de que debemos extender retroactivamente la salvaguardia lockeana hasta abarcar toda la historia pasada es suponer que la moralidad es natural en un sentido de "natural" más fuerte del que Gauthier puede aceptar. Porque, al fin y al cabo, él defiende la tesis de que la moral es 'por acuerdo'" (art. cit., p. 144).

diferentes. Incluso cabe admitir que la idea de una salvaguardia limitadora de la posición inicial de negociación sea un requisito racional para el constructivismo moral, mas no así para el político. Desde ese punto de vista, ambas teorías serían correctas, cada una en su nivel.

Ahora bien, en la medida en que el contractualismo político hobbesiano conlleva un convencionalismo moral, la crítica de Narveson a Gauthier tiene sentido, precisamente porque la moral por acuerdo *no* es una teoría convencionalista. Ahora bien, tiene sentido sólo desde el punto de vista (político) de Buchanan-Narveson. Desde el punto de vista del contractualismo moral, la crítica es irrelevante porque, en contra de lo que cree Narveson, un contractualista sí puede aceptar que la moralidad es "natural" en sentido fuerte; en el sentido de que una relación causal necesaria (natural, si se quiere) liga a cualquier ser racional, puesto en una situación de interacción, con ella. Tal vez sea imposible compatibilizar un contractualismo moral (que incorpora, de alguna forma, una teoría política contractual ya modificada respecto al contractualismo hobbesiano puro) con una teoría político-económica como la de Buchanan[198]. Pero ello no será debido a la distinta concepción de los límites a la interacción natural, ya que ésta diferencia proviene de las distintas necesidades impuestas por los objetivos de ambas teorías. En este sentido, ambas opciones son válidas y ninguna ofrece base para la crítica de la otra.

También puede ponerse en relación con el distinto alcance (o doble nivel) de las teorías una segunda razón que serviría para descartar la crítica de Narveson. Se podría argüir que la inadmisión de límites a la interacción natural

[198] No hemos profundizado lo suficiente como para afirmarlo, pero tal compatibilidad parece rigurosamente imposible, ya que la cooperación *exige* una estructura de derechos (determinados por la salvaguardia) que el contrato constitucional de Buchanan no habría respetado. Desde esta óptica, el sistema de cooperación nacido del pacto hobbesiano sería injusto desde el punto de vista de la moral por acuerdo y, por tanto, incompatible con ella. Ahora bien, esta rigurosa incompatibilidad teórica podría relajarse en la práctica si se acepta que lo política y económicamente admisible y útil o lo jurídicamente legítimo no siempre coincide con lo moralmente correcto. El contractualismo político hobbesiano de Buchanan serviría como *mínimo* de corrección política y de justicia distributiva en el mercado; el contractualismo moral de Gauthier serviría como criterio *mínimo* de actuación para las instituciones cooperativas y de orientación para las reformas políticas y jurídicas.

significa que Buchanan no toma en serio el postulado de igual racionalidad[199]. Aunque las capacidades naturales sean moderadamente desiguales, y eso conduzca a posiciones iniciales más o menos ventajosas derivadas de la predación, la igual racionalidad de las partes implica que el menos aventajado reclamará que el aprovechado renuncie a la parte de su utilidad derivada exclusivamente de la predación, y éste reconocerá la necesidad de tal renuncia para abrir la puerta de la cooperación, que le beneficia a él más incluso que al desaventajado (porque, debido a su mayor capacidad o "inversión", espera obtener más de la aplicación del principio de concesión relativa *minimax*). La sumisión a los pactos que perpetúan una distribución natural basada en la predación sólo demuestra una falla en la racionalidad de una de las partes. Esto es evidente en el texto de Narveson que reproducíamos arriba: el hecho de que uno esté mejor de lo que estaba no es, por sí sólo, una razón convincente para asentir al pacto.

En definitiva, no puede decirse que las lecturas de la salvaguardia analizadas hasta ahora excluyan la posibilidad de justificar su papel en la teoría del contrato moral. Aunque sí ponen de manifiesto las ambigüedades de Gauthier, que no define con nitidez los perfiles contractuales de este elemento de su teoría. Afortunadamente, existen contribuciones que han tratado de completar esa ambigüedad de Gauthier de acuerdo con su teoría. De la mano de una de ellas intentaremos nosotros también dar un sentido más preciso a la salvaguardia.

c) Las dimensiones de la salvaguardia.-

Dedicaremos este epígrafe principalmente a debatir la propuesta de P.

[199] Aunque, en su teoría del pacto constitucional, explícitamente circunscribe las desigualdades naturales a los gustos o preferencias, las capacidades o talentos y el medio; mientras, la racionalidad se supone implícitamente postulada como igual (Cfr. Buchanan, *The Limits of Liberty*, cit., pp. 54-55).

Danielson: en su ensayo "The Lockean Proviso"[200], Danielson avanza una lectura de la salvaguardia que, pese a ser replicada por Gauthier[201], sólo pretende llevar a sus últimas consecuencias el argumento de *MA*.

En principio, Danielson acepta el carácter "transcendental" de la salvaguardia: es una "condición de posibilidad" del contrato. Acepta también la parte de su contenido que se refiere a la interacción directa —esto es, la prohibición de la coacción— entre los agentes naturales. Pero niega que los derechos de propiedad adquiridos en el estado de naturaleza deban configurar la posición inicial y, en este sentido, rechaza el contenido de la salvaguardia referido a estos derechos[202]. Los derechos de propiedad (y la distribución de los mismos) son fruto del pacto, y han de estar subordinados al principio de beneficio mutuo *maximin*[203]. El resultado de esta concepción es justamente el opuesto al de Buchanan: un estado de naturaleza restringido por una salvaguardia que *prohíbe* adquirir cualquier propiedad (sea fruto de la predación o de la interacción permitida por la salvaguardia de Gauthier). Así, nadie lleva a la mesa de negociación, como dotación inicial, nada más que sus capacidades y talentos (sus derechos personales). Como consecuencia, la distribución post-contractual es más igualitaria que en las teorías de Buchanan y Gauthier.

---

[200] en Vallentyne, P. (ed.), *Contractarianism and Rational Choice*, cit., pp. 99-111; es re-edición de una parte de Danielson, P., "The Visible Hand of Morality", cit.

[201] Cfr. "Moral Artifice", cit., pp. 406-413.

[202] Lateralmente, hay que anotar lo acertado de la distinción de Danielson en el sentido siguiente: el interés en justificar la propiedad privada (y, sobre todo, el acto de "apropiación") es característico de las teorías de los derechos (Locke, Nozick); pero debería ser, en principio, ajeno a una teoría del contrato (Rousseau, Hobbes, Rawls). La salvaguardia cumple, en las teorías de los derechos, la función de limitar la extensión de las apropiaciones particulares (es, literalmente, una "ley natural" —sea en forma de origen divino, sea en forma de "restricción moral indirecta"— que se impone, no ala interacción natural en cuanto tal, sino a los concretos actos de apropiación, que no tienen por qué ser interacciones). En las teorías del contrato, la extensión de la propiedad y sus límites puede muy bien fijarse contractualmente.

Pero si la distinción es acertada, no menos acertado es captar que el problema de Gauthier consiste precisamente en que trata de compatibilizar ambas teorías: justificar contractualmente las instituciones cooperativas, pero mantener una base individualista —aunque sea mediante la "transcendentalización" de los derechos individuales de propiedad— de las mismas que garantice la conformidad de cada persona.

[203] En el mismo sentido, puede verse la interpretación del derecho de herencia realizada por Kevin Sauvé, art. cit., p. 173.

El énfasis en la fuente contractual de los derechos de propiedad es un acierto de Danielson. Aunque quizá ese énfasis le lleva a "naturalizar" los derechos personales de un modo contrario a su intención. Este peligro —olvidar el carácter "transcendental" de (todos) los derechos— parece estar presente en la mayoría de los críticos tanto como en Gauthier. Pero el argumento de Danielson sobre los derechos de propiedad nos pone (a nosotros y a Gauthier) sobre la pista del verdadero sentido de la salvaguardia.

En efecto, en su réplica a Danielson, Gauthier se ve compelido a explicar cómo, incluso en ese nivel "transcendental", se pueden justificar derechos de propiedad naturales, en la medida en que su existencia mejora las perspectivas de cooperación de todos los agentes y la posición relativa de algunos de ellos. Los derechos de propiedad naturales (limitados por la salvaguardia) permiten que la posición inicial de negociación sea pareto-superior a la situación de no-interacción. Y, dentro de los posibles resultados óptimos de Pareto, es racional que cada agente demande que los restantes negociadores reconozcan como su derecho exclusivo el resultado natural que él obtuvo ya que, en la medida en que ese resultado se da en una situación óptima y restringida por la salvaguardia, ninguno puede objetar haber sido perjudicado por la eventual apropiación anterior que produjo el resultado natural.

Dicho de otra forma: en la negociación, la dotación inicial desde la que cada agente negocia no se compondrá sólo de los derechos personales, capacidades y talentos, sino que cada uno reclamará que esa dotación incorpore también lo que habría obtenido en un estado de naturaleza sin violar la salvaguardia, incluidas las posesiones materiales. Los demás atenderán esta reclamación —pero no una reclamación mayor[204]— y reconocerán un derecho sobre esas apropiaciones.

El límite que debe fijar la salvaguardia para las apropiaciones en el estado de naturaleza está sujeto a discusión. Tal vez este debería ser muy estrecho, como sostiene Danielson, o infinitamente laxo, como defendería Buchanan. Pero lo cierto es que, de acuerdo con los postulados del contractua-

[204] Cfr. Gauthier, D., *MA*, p. 227.

lismo moral liberal, Gauthier parece autorizado a defender una salvaguardia cuya extensión queda casi a medio camino entre la libertad absoluta del estado natural hobbesiano y las restricciones que impedirían toda apropiación.

d) Conclusión: la salvaguardia y la cooperación.-

El concepto de la salvaguardia aproxima a Gauthier a la tradición del derecho natural y, por eso, lo convierte en blanco del tipo de críticas normalmente dirigidas contra esa tradición. Se podrían encontrar otros ejemplos de objeciones contra la salvaguardia, pero la orientación de la mayoría de ellos sería semejante a la de los que hemos elegido. La teoría de los derechos distorsiona la percepción de la salvaguardia, porque se trata de una teoría política o jurídica, interesada en justificar moralmente un área de libertad individual frente al poder estatal. Su interés es la relación individuo-estado; la moral es sólo su instrumento. La de Gauthier es una teoría moral: su interés es justificar ante el individuo restricciones morales; la justicia (como imparcialidad) y las instituciones políticas sirven a esa justificación, porque sólo en un marco de cooperación mutuamente beneficioso cabe esperar que un individuo racional acepte límites a su comportamiento maximizador. Esta diferencia de enfoque es crucial para comprender el papel de la salvaguardia.

En el marco del proyecto de una moral por acuerdo, la salvaguardia cumple un papel primordial en el engranaje global de la teoría. Ahora bien, tratar de entender ese papel en el estrecho perímetro de un estado de naturaleza "clásico" es vano.

Como hemos visto, el contractualismo moral liberal no describe un estado de naturaleza al estilo clásico (como hacen Nozick o Buchanan), sino que contempla en abstracto la estructura de la interacción no-cooperativa (que podemos llamar "natural" sólo metafóricamente) entre agentes estrictamente definidos como perfectos maximizadores racionales. A partir de esa estructura surge la necesidad de la cooperación; ésta depende del pacto y su posterior cumplimiento. Todos los elementos de la teoría se subordinan al objetivo final

de que pueda exigirse justificadamente el cumplimiento de una estrategia conjunta a un individuo racional. Todo en la teoría está pensado para responder al escéptico, al egoísta, al explotador. Por eso la salvaguardia "representa el límite menor racionalmente aceptable por personas que quieren evitar una interacción costosa con otros, y el límite mayor racionalmente aceptable por personas que quieren ser libres para obtener el mayor beneficio posible"[205]. La salvaguardia representa el compromiso necesario para que la negociación sea aceptable para agentes igualmente racionales, pero auto-interesados y moderadamente desiguales en sus capacidades.

A la vez, la salvaguardia explica, por referencia al contrato, la existencia de derechos individuales y derechos de propiedad, y justifica su mantenimiento sin necesidad de apelar a tradiciones o a "derechos individuales naturales" (en el sentido de Nozick). Porque, aunque se denominen apropiadamente "derechos naturales" dado que han de suponerse existentes en el estado previo al contrato, se trata de derechos originados y legitimados *en* el contrato mismo. Al igual que la racionalidad de las partes conduce a un principio racional de negociación, lleva también necesariamente a una determinada estructura de derechos que configura la posición inicial de negociación.

Desde nuestro punto de vista, habría sido tal vez menos equívoco dar a las dotaciones iniciales permitidas por la salvaguardia un nombre nuevo, no el de "derechos", y reservar éste para la estructura de obligaciones y propiedades reconocidas en el sistema de cooperación y mercado post-contractual. Pero, en este punto, Gauthier ha querido contribuir a la teoría de los derechos naturales mostrando una conexión —tal vez inesperada— entre el contractualismo y dicha teoría.

Otros contractualistas han prescindido de esa conexión, al concebir los derechos inequívocamente como resultado del pacto. Pero esa concepción conlleva la admisión de una posición inicial de negociación fruto de una interacción no-cooperativa eventualmente coactiva o predatoria. Buchanan ha demostrado que tal posición inicial podría dar lugar a un contrato constitucional mutuamente beneficioso, base para ulteriores pactos post-constitucionales que

[205] Gauthier, D., *MA*, p. 227.

establecerían estructuras de intercambio y cooperación cuyos resultados se aproximarían asintóticamente a los predichos por el principio de concesión *minimax*. Pero Buchanan muestra que es racional cumplir con el pacto constitucional sólo a base de relajar la suposición de igual racionalidad, adoptando premisas más "empíricas". Si se mantiene el postulado de igual y perfecta racionalidad, como Gauthier hace, la teoría de Buchanan es incapaz de demostrar que sea racional conformarse a pactos que implican transferencias improductivas de utilidad entre los miembros de la sociedad.

Si la cooperación es mutuamente beneficiosa, entonces los mejor dotados para la predación y la defensa abandonarán sus prácticas (o renunciarán a reclamar derecho alguno sobre lo adquirido mediante ellas) para ponerse en situación de negociar y pactar para hacerla posible. Las victimas de su habilidad agresiva no aceptarían un pacto menos beneficioso para ellas. Mejor dicho, lo aceptarían sólo como medio para escapar de las prácticas predatorias, pero, una vez eliminada la amenaza o la coacción, reclamarían tanta parte del beneficio cooperativo como les correspondería si el contrato se hubiera desarrollado desde una posición inicial no coactiva e imparcial. Es decir, un contrato que no corrija, desde su origen, las adquisiciones fruto de la predación y la violencia, es inestable. Se puede mantener como contrato social o político, dado que se implementan mecanismos coactivos a ese objeto. Pero sería irracional como pacto moral, pues carecería de capacidad motivadora ante el individuo. Si se supone que los agentes son igualmente racionales, sólo un contrato cuyo origen pueda concebirse, al menos, como un estado natural atenuado por la salvaguardia, podrá reclamar la categoría de contrato moral.

Así, la salvaguardia establece un límite hipotético a la interacción natural y con ello permite que la cooperación sea racional para cada individuo, porque sus frutos se dividen de acuerdo con un principio de distribución que tiene en cuenta los intereses de cada uno y desde una situación inicial imparcial (en la cual todos gozan de iguales oportunidades). Como se demostró en el punto anterior, es racional estar dispuesto a cumplir este tipo de acuerdos siempre que otros lo hagan. Se ha mostrado, por tanto, la posibilidad de la cooperación racional y, por añadidura, se ha mostrado que la cooperación implica la

aceptación de restricciones imparciales a la libre persecución del auto-interés. Gauthier sostiene que estas restricciones pueden identificarse con la moralidad, de modo que considera cumplido su objetivo de derivar la moralidad a partir de las débiles premisas de la racionalidad como maximización. Para valorar la plausibilidad de esta afirmación quizá convenga tomar un respiro y recapitular el argumento desarrollado en los puntos anteriores.

6. Moral contractual e individuo: recapitulación final

Los elementos de la teoría de Gauthier reflejan los componentes clásicos de las teorías del contrato social: un estado de naturaleza, un pacto o convenio, una estructura institucional fruto del pacto que ha de ser mantenida (coactivamente en ocasiones) y que está legitimada por su origen en la voluntad unánime de los miembros de la sociedad. Una teoría contractual de la moralidad —que debe prescindir, por definición, de la coacción como método de implementar los acuerdos— ha de explicar, además, la racionalidad de cumplir los pactos concertados (o las promesas hechas, que viene a ser lo mismo); y esta explicación pivota sobre la imparcialidad o "limpieza" de la negociación y el pacto.

El concepto central de una teoría contractual es el "acuerdo" —lo cual es uno de los puntos de coincidencia entre contractualismo y convencionalismo. Pero "acuerdo" tiene dos sentidos, en el marco del contractualismo moral:

Por un lado, un acuerdo se entiende como una *negociación* en la que participan los agentes auto-interesados para, intentado cada uno maximizar su beneficio, adoptar un convenio (estrategia conjunta cooperativa) que garantice un resultado colectivo óptimo. El final de la negociación será, por tanto, una decisión unánime; aunque implique cesiones y haya estado antecedida por regateos y transacciones.

Por otro lado, un acuerdo se entiende como el *contrato* por el cual se asegura (coactivamente o no) el cumplimiento del convenio pactado, y al cual se adhieren voluntariamente las personas por su propio interés.

En sus dos sentidos, el concepto de acuerdo incorpora las nociones básicas del contractualismo: libertad y racionalidad individuales, respeto a los intereses y preferencias de cada persona, unanimidad en el pacto fundante, cumplimiento voluntario de lo pactado.

La teoría de Gauthier articula estas nociones de un modo insuperable en las conocidas sub-teorías de la negociación racional, del cumplimiento, y de la posición inicial. Además, estas teorías permanecen fieles al principio de radical minimalismo en las premisas. Por separado, cada una de estas sub-teorías es suficientemente rica como para ver en ella un argumento capaz de mostrar la racionalidad de la moralidad:

Basándose en la teoría de la negociación, se podría afirmar que es racional adoptar los principios de justicia que pudieran haber sido pactados por agentes racionales tras una negociación ideal llevada a cabo desde una convenientemente definida posición originaria. Se configuraría, así, una teoría de la justicia paralela a la de Rawls, con la diferencia de incorporar la negociación como mecanismo de decisión "tras el velo de ignorancia"[206].

Basándose en la teoría del cumplimiento y en el desarrollo de la Teoría de la Decisión Racional que supone la idea de la maximización restringida, podría afirmarse que la moralidad consiste en adoptar un tipo de racionalidad especial, que capacita al individuo para cumplir promesas y pactos. No es tanto que racionalidad cooperativa dé lugar a la moralidad al hacer racional el cumplimiento de acuerdos racionales y justos; es que ella misma tiene la consideración de una disposición moral, en cuanto supone una transformación del individuo directamente maximizador en una persona dispuesta a la cooperación. Esa transformación revela el carácter moral de los individuos, y —teniendo en cuenta que la maximización restringida se justifica sobre bases de maximización directa— es prueba suficiente para afirmar, siguiendo a Baier, que la moralidad es el verdadero interés del agente racional.

Por último, sobre la base de la teoría de la posición inicial, podría argüirse que la moralidad se cifra en el respeto a ciertos derechos individuales naturales. No sería una fundamentación caprichosa. La moral consistiría en una serie de restricciones necesarias (exclusivamente las imprescindibles para garantizar los derechos protegidos por la salvaguardia) para hacer posible la cooperación social y, por ende, el beneficio mutuo. Por tanto, también desde

---

[206] Esta es, por otro lado, la visión que de su proyecto tiene Gauthier durante la década de 1975 a 1985; el tiempo que va desde la publicación de "Justice and Natural Endowment: Toward a Critique of Rawls's Ideological Framework" hasta la de "Bargaining and Justice".

esta perspectiva la moralidad tendría como fundamento el interés individual, sin necesidad de apelar a un origen distinto de los derechos.

Pues bien, de estas posibles lecturas, ninguna se ajustaría a una teoría moral contractual; ninguna capta el argumento contractualista como tal. Si hemos logrado explicar acertadamente el sentido de una moral por acuerdo, debe ser evidente que no se trata de una moral basada en los derechos naturales, ni una re-edición modificada de la *Teoría de la Justicia*, ni un refinamiento del "egoísmo ilustrado" basado en una teoría psicológica o antropológica sobre la motivación o el carácter moral. Los tres núcleos teóricos de la moral contractual forman parte de un engranaje singular que, convenientemente acoplado, sirve para mostrar (a) el modo en que la acción moral (un tipo de acción directamente contraria al auto-interés del agente en muchas ocasiones) puede justificarse ante el individuo y (b) cómo puede deducirse un criterio intersubjetivo de justicia para las instituciones cooperativas. Ambas funciones están interconectadas, y ambas exigen una comprensión unificada de la teoría. En algún momento hemos utilizado el símil de una cadena circular cuyo último eslabón engarza con el primero. No es un símil acertado. No hay sucesión, ni siquiera circular, en la teoría. Funciona como un todo, como una unidad; ninguna parte es "anterior", ni temporal, ni lógica, ni epistemológicamente, a las otras. En términos kantianos —empleados únicamente como metáfora explicativa— se podría decir que el contrato moral explicita las condiciones de posibilidad de cada acción moral. Pero esas condiciones ni son anteriores al acto moral, ni pueden ser conocidas por nosotros con independencia del mismo, ni se pueden "ordenar" como un razonamiento deductivo.

Cada uno de los elementos de la teoría implica los demás y está implicado por ellos: es racional negociar y pactar una estrategia conjunta si y sólo si la negociación es limpia e imparcial y se espera que va a ser cumplida; es racional conformarse a (y cumplir) una estrategia conjunta si y sólo si es fruto de una negociación imparcial; el resultado de una negociación racional es imparcial si y sólo si la situación inicial lo es; etc.

El contrato moral representa la versión más abstracta de contractualismo: el carácter narrativo propio de la tradición, que incluía en ocasiones hasta

ejemplos históricos, es completamente sustituido por la exposición de las entrañas lógicas del razonamiento que hace individualmente plausible el contractualismo como esquema normativo de justificación. El complejo argumento se dirige al escéptico; mostrando que la realización de *ese* acto moral concreto es directamente interesante para él[207]. Y esa demostración no apela a un pasado histórico o heurístico, sino a la lógica de la interacción misma y a los supuestos (transcendentales, si se quiere) en ella incorporados.

Pero justificar no quiere decir motivar. Como individuos, podemos ver justificada una acción o una institución y, sin embargo, no realizarla o no aceptar sus reglas. La flaqueza de la voluntad explicaba tradicionalmente este problema —un problema que ninguna teoría normativa ha logrado solucionar completamente. Gauthier apela, como ya sabemos, a la maximización restringida para explicar cómo, en tanto que seres racionales, hemos de sentirnos motivados a cumplir con las normas morales, en la medida en que reflejen adecuadamente el principio de una negociación racional y la estructura de derechos establecida por la salvaguardia.

La maximización restringida tiene así —como, por cierto, las demás nociones nucleares de la teoría— un papel *interno* al argumento y otro externo. Su papel interno consiste en hacer racional la negociación al garantizar el cumplimiento del pacto, y hacerla justa al garantizarlo exclusivamente bajo ciertas condiciones. Su papel externo está en relación con la motivación: la fidelidad a los mandatos de la moralidad depende de que, de hecho, podamos

[207] Si es capaz de razonar correctamente que se trata de un acto ordenado por una regla reguladora de cierta práctica cooperativa que habría sido acordada *por él mismo* de haberse encontrado en una posición inicial *anterior a la cooperación* negociando con sus semejantes una mejor gestión de los recursos a fin de alcanzar un óptimo social lo más beneficioso posible para él mismo. Si reconoce, simultáneamente, que la posibilidad de tal contrato está supeditada a su cumplimiento y ésta, a su vez, a un reconocimiento mutuo de ciertos derechos inalienables. Si razona y reconoce todo esto, entonces justificaría, ante sí mismo como agente racional, el acto moral que realiza (Pero este razonamiento es bastante improbable, a no ser que el agente en cuestión sea un filósofo moral; porque, entre las instituciones cooperativas se encontrará seguramente un mecanismo de socialización que evita que este tipo de justificación sean habitualmente necesarias).

ser concebidos como maximizadores restringidos[208].

En este punto, hemos de abandonar el postulado individualista tal como fue formulado en el capítulo segundo y tratar de ver qué tipo de individuo aceptaría y cumpliría una moral por acuerdo, y si ese tipo de individuo se asemeja a las personas reales. Este movimiento no deja de ser cuestionable y tentativo, en la medida en que no somos investigadores sociales, y no podemos reclamar para nuestras afirmaciones nada más que su plausibilidad desde el punto de vista del sentido común.

Es evidente que las personas reales no somos como el individuo racional y auto-interesado que postula el liberalismo. Nos atan lazos afectivos de muchas clases que conforman un denso entramado de "intereses en los intereses ajenos" completamente opuesto a la postulada independencia individual. Se podría cuestionar qué poder ejercerá sobre nosotros, personas reales, una teoría tan abstracta y construida desde premisas tan relativamente distantes de nuestra realidad. De hecho, quienes confían en la afectividad como fuente de la moralidad no sólo cuestionan, sino que niegan que este tipo de teorías pueda tener alguna relevancia para "la trama moral" de nuestra vida cotidiana[209]. No obstante, ya señalábamos en su lugar, que los caracteres del postulado individualista están extraídos (abstraídos, habría que decir) a partir del perfil medio del hombre económico que mayoritariamente puebla nuestras sociedades liberales. En cada uno de nosotros y nosotras hay necesariamente una parte de maximizador, o al menos de satisfactor, directo —la que concierne a la estructura básica de nuestras necesidades biológicas. Nuestra racionalidad es, inicial y primariamente, instrumental. El paso adelante de la moral por acuerdo, respecto al contractualismo hobbesiano, consiste precisamente en reconocer que estos postulados y descripciones corresponden quizá al núcleo (biológico) de lo

---

[208] Se puede adivinar que la negociación tiene un papel interno consistente en servir como mecanimso de selección de una estrategia conjunta, y otro externo, alproveer a la sociedad de unprincipio de justicia que opera en las instituciones; la salvaguardia garantiza, internamente, una posición inicial de negociación racional, externamente, sirve para reconocer una estructura básica de derechos.

[209] Un ejemplo de este tipo de postura es el artículo de A. Baier, "Pilgrim's Progress", *Canadian Journal of Philosophy*, vol. 18, nº 2, junio 1988, pp. 315-330.

humano, pero no caracterizan la *totalidad* de lo humano, más bien al contrario, constituyen en todo caso un núcleo muy pequeño.

El contractualismo moral muestra algo que, sin embargo, no es evidente: que no somos máquinas de maximizar. Mostrar esto, y no simplemente suponerlo, requería partir de los estrictos postulados sobre el individuo y la racionalidad. Pero, aceptando aquellos postulados se alcanza la posibilidad de un individuo —el maximizador restringido— que transforma su paradigma de racionalidad para hacer racional el cumplimiento de los acuerdos[210]. Si el argumento de Gauthier es convincente, se ha mostrado cómo una disposición moral surge a partir del mero auto-interés. Con ello, hay que suponer que la capacidad moral está ya de alguna manera incorporada en la racionalidad humana, incluso en su nivel más elemental.

Pero como consecuencia de esa "transformación moral", los individuos en una sociedad liberal, aun siendo independientes y reservando muchas parcelas de su interacción a la competencia y al auto-interés, son agentes dispuestos a la cooperación con sus semejantes. La cooperación amplía por sí misma las expectativas de beneficio mutuo[211] y, cabe esperar que, a través de la valoración de la cooperación, los individuos lleguen a valorar también a sus compañeros cooperadores[212]. Así, asociamos al individuo liberal con una persona libre en sus afectos y en sus creencias, pero capaz de aceptar sinceramente las normas que le vinculan a la comunidad de la que es parte y de la

---

[210] Esto es enfatizado por Gauthier en muchos textos. Como ejemplo valga en siguiente fragmento de "Moral Artifice" (p. 416): "Cuando comenzamos a actuar basándonos en nuestras *representaciones, tal vez somos maximizadores naturales; pero nuestra capacidad de representarnos* estados de cosas y de reflexionar sobre nuestras representaciones nos permite tratar el principio de maximización como algo tan cuestionable como cualquier otra cosa. No sólo el contenido de la preferencia humana, sino también su forma, carece de rasgos fijos."

[211] Obsérvese que hasta aquí nos hemos centrado en el valor *instrumental* de la cooperación en relación con los intereses de cada individuo; pero la cooperación produce un beneficio añadido: la posibilidad de realizar actividades intrínsecamente cooperativas, es decir, compartidas (tales como cantar a cuatro voces, construir una pirámide o formar un equipo de fútbol). En este sentido, *la cooperación no sólo sirve a los intereses individuales, sino que expande el horizonte de* posibilidades de los individuos y los grupos, haciendo posibles satisfacciones rigurosamente imposibles para el individuo solitario.

[212] Cfr. Gauthier, D., *MA*, p. 336.

que, en buena medida (y él es consciente de ello), nacen las posibilidades que le permiten desarrollar su libertad y sus capacidades[213]. En el nivel moral, esta comunidad tiende a identificarse con la especie, y sus normas con las normas mínimas de una moral universal basada en el beneficio mutuo, pero que lo transciende gracias a la afectividad libre desarrollada por los individuos (sobre esto volveremos más abajo).

Reconocer la inserción social del individuo liberal y su capacidad afectiva no significa, sin embargo, que pretendamos trastocar a última hora el alcance de la teoría. Una moral por acuerdo es, si se quiere, una teoría moral elemental. Con toda su complejidad, lo que muestra es bien poco: por un lado, que la moralidad se puede justificar incluso desde presupuestos de racionalidad mínimos y, por tanto, que cabe afirmar en un sentido ampliamente aceptable que ciertas restricciones morales son racionales; por otro, que el criterio para determinar qué restricciones son racionales es un principio de beneficio relativo *maximin*. La moral por acuerdo muestra tan sólo las bases instrumentales de la moralidad. Desde luego que nuestras convicciones morales no se sostienen sólo apelando a su utilidad[214], pero es importante haber demostrado que *también* pueden defenderse sobre esa base tan simple.

Las restricciones que se asociarían a esta moralidad son, probablemente, mucho menos exigentes que las de la mayoría de nuestros sistemas éticos heredados. Pero serían las que *incluso* un individuo económico aceptaría. Si algunos de nosotros, individuos liberales, estamos dispuestos a complementar esas restricciones con otras derivadas de nuestros afectos o nuestro altruismo, miel sobre hojuelas. En este caso, la moral por acuerdo no se dirige precisamente a nosotros.

---

[213] En este punto, la visión de una sociedad liberal (y creemos que es una visión compartida por autores como Gauthier y Rawls) es mucho menos negativa que aquella que refleja Tönnies cuando conceptualiza la *Gesellschaft*. Frente a la visión de hostilidad hacia la sociedad y la tensión entre individuo y sociedad, las teorías liberales de la justicia y la moralidad han encontrado un nexo entre sociedad e individuo que re-sitúa al agente moral en su lugar natural: la ciudad (Cfr. Helena Béjar, *El ámbito íntimo*, Madrid, Alianza, 1988, pp. 90 y ss.).

[214] Cfr., sobre esto, la interesante conclusión de Den Hartogh, "The Rationality of Conditional Cooperation", cit., p. 421.

Pedir que una teoría moral post-ilustrada dé cuenta de la moralidad asociada a (o derivada de) las relaciones afectivas de los humanos sería pedir demasiado. Si una teoría moral es capaz de mostrar (a) que es racional para cada individuo *aceptar* sus reglas como fruto de un acuerdo con sus semejantes; (b) que es racional para cada individuo *actuar* conforme a sus reglas; y (c) que el resultado, si todos siguen las normas, es *óptimo*, entonces, como dice Gauthier "¿quién puede pedir más?"[215]. Esa teoría habría pasado las pruebas que, desde el punto de vista de la racionalidad, pueden presentársele.

[215] Cfr. "Moral Artifice", cit., p. 389.

Capítulo V

Conclusión:
Razón pública y moralidad

1. La moral de una sociedad liberal

El contractualismo es, como hemos repetido en varios lugares, una respuesta al escéptico moral. Los argumentos convencionalistas o contractualistas son los únicos aceptables para un individuo emancipado, autónomo, en un mundo "normativamente desencantado". En esta situación, la instancia normativa, o razón común, ha de ser *construida* por las racionalidades individuales, y con los materiales aportados por éstas. Esa construcción tiene diversos niveles: el nivel del consenso político, el nivel del acuerdo constitucional y de los acuerdos post-constitucionales, el nivel de la discusión pública de los principios, etc. El objetivo del contractualismo moral es hacerse cargo del nivel más elevado de esa construcción intersubjetiva, que consiste en la reconstrucción ideal del acuerdo racional hipotético en que se funda la posibilidad misma de la sociedad y de la discusión política pública. Las instituciones sociales, y la propia sociedad, se basan en el pre-supuesto de un acuerdo moral: el acuerdo que explicita, cláusula por cláusula, las condiciones de la cooperación, sus límites, su justificación, su necesidad. El acuerdo moral trae a la luz la condición de posibilidad de la ética pública, pues trata de explicar por qué *cada uno* de nosotros es un animal político, en vez de simplemente suponerlo. Como consecuencia, se justifica racionalmente un esquema de cooperación que

podemos identificar con la "sociedad liberal": una comunidad de personas libres, autónomas, que comparten un ideal racional común como núcleo intersubjetivo de normatividad.

Cabe objetar que el postulado individualista que se emplea como premisa del contrato, ya pre-supone un tipo de sociedad-asociación, revestida de las características de la sociedad liberal contemporánea. En realidad —arguye esta crítica— la conclusión era previsible, dadas unas premisas que ya la pre-contenían de algún modo.

Respondimos a esta objeción en el capítulo II, argumentando que un postulado no pre-supone, por definición, ninguna conclusión determinada y que, por lo que se refiere a la racionalidad, el concepto instrumental aceptado es tan elemental que difícilmente puede negarse que, al menos en parte, representa un buen fragmento de racionalidad humana, en cuanto somos seres orientados a fines.

Lo que la teoría consigue a partir de tan mínimas premisas excede con mucho las expectativas que las mismas podrían sugerir[1]. El individualismo radical y la racionalidad instrumental funcionan como presunciones en contra de la moralidad y de la comunidad. La previsión más optimista anticiparía, como mucho, una asociación de maximizadores a largo plazo, respetuosos de una moralidad convencional al servicio de sus intereses particulares. El mérito del contractualismo moral consiste en mostrar que es racional ser moral, no sólo instrumentalmente —aunque resulta haber también una justificación instrumental de la moralidad—, sino "esencialmente"[2]. Una moralidad meramente instrumental sería auto-contradictoria; el contrato moral desvela, en el concepto de maximización restringida, una auténtica "transformación moral". Con ello, las presunciones en contra de la moralidad se auto-cancelan. Contra todo pronóstico, la moralidad surge de la mera racionalidad y asistimos al

[1] Así, al comienzo del desarrollo de su teoría, Gauthier mismo no cree que la justificación que se puede ofrecer de la cooperación sea algo más que una "moral instrumental del hombre económico". Véase, p. ej. "Rational Cooperation", *Noûs*, 8 (1974), pp. 53-65; p. 62.

[2] Empleamos este término en el sentido en que lo hace Gauthier en el capítulo final de *MA*, al referirse a una sociedad "esencialmente" justa y a la justicia "esencial", frente a una hipotética sociedad "instrumentalmente" justa o una justicia "instrumental".

nacimiento de una sociedad justa como unión voluntaria de agentes morales.

Una sociedad justa no es, por tanto, una asociación de individuos enlazados por un contrato coactivamente impuesto; ni la moralidad es la simple y astuta "maximización a largo plazo". Una sociedad justa aparece, para el contractualista liberal, como una unión de agentes morales que comparten un núcleo de convicciones comunes que puede ser reconstruido como un acuerdo racional para el beneficio mutuo.

El contractualismo muestra que la disposición moral supera el test más severo: el que conjuntamente imponen la eficiencia, el auto-interés y el escepticismo. Se demuestra así que el agente racional es necesariamente ya un agente moral, que no cabe racionalidad sin disposición a la moralidad, que elegir moralmente *es* elegir racionalmente.

La racionalidad de la moralidad garantiza la adherencia al principio de justicia producto del contrato moral. Este principio configura una sociedad en la que cada individuo es libre para orientar su vida conforme a valores subjetivos que pueden divergir de los de sus semejantes; y aún se ve animado a hacerlo, porque en una sociedad liberal el pluralismo y la diversidad llegan a ser valorados por sí mismos. Donde el beneficio mutuo se reconoce como producto de la diversidad de los talentos, capacidades, preferencias e intereses de los demás, es lógico que esta diversidad se fomente y se valore. La emancipación individual y la libre afectividad garantizarán, en una sociedad regida por el principio de concesión *minimax*, la adhesión voluntaria a las instituciones cooperativas. Por otro lado, estas instituciones, además de servir al beneficio mutuo, expanden —ya lo anticipábamos al final del capítulo anterior— el horizonte de posibilidades de las personas, de modo que es plausible suponer que, entre las preferencias de los individuos liberales se asiente una preferencia por la sociedad como sistema cooperativo; una preferencia independiente del hecho de que la cooperación sea el mejor mecanismo para el beneficio mutuo (entendido como beneficio de cada uno). Así, los miembros de una sociedad liberal están interesados en las instituciones cooperativas, y en la justicia que les es inherente, por sí mismas.

De este modo, la normatividad mínima que fija el marco de la cooperación no es tan vacía y tan formal como pudiera parecer, porque, al reflejar las

preferencias de cada agente, se inclinará afirmativamente hacia ciertas instituciones que engrandecen las expectativas de vida de las personas, aunque no vengan estrictamente requeridas por el principio *minimax*. Gracias a estas instituciones podemos entender que la sociedad liberal es una sociedad donde la moralidad abarca un ámbito mucho mayor que el beneficio mutuo.

La moralidad instrumental del hombre económico que Gauthier entreveía en 1974 no es tal cosa; sino más bien una disposición moral esencial que constituye al sujeto liberal y que sienta las bases de la gran comunidad en sentido moral.

2. Contrato moral, universalismo y racionalidad

La sociedad liberal se asienta, como decíamos, sobre el acuerdo hipotético que explicita las condiciones de posibilidad de la cooperación. El contractualismo moral sostiene que ese acuerdo hipotético constituye, no sólo el pre-supuesto de una sociedad concreta, sino el fundamento de una moral universal. Tal acuerdo es la forma de un argumento cuya premisa mayor es una concepción de la racionalidad prudencial ampliamente aceptada, desde la que se justifica, mediante una serie de razonamientos heurísticos válidos, que bajo condiciones generalmente prevalentes y plausibles es *racional* aceptar restricciones morales (y no simplemente someterse a ciertas convenciones). El argumento muestra que cualquier agente racional situado en las "condiciones de la justicia", tiene un motivo prudencial para adoptar una disposición moral.

Por nuestra parte sostenemos que, pese a sus imperfecciones, y pese a las reformulaciones a las que ha sido sometido —y a otras futuras que, sin duda, exigen sus deficiencias—, el argumento de Gauthier constituye la única teoría ética que puede arrogarse el nombre de contractualismo moral en sentido estricto, según nuestra definición del capítulo tercero.

El contractualismo moral se distingue del convencionalismo porque, partiendo del subjetivismo axiológico y del individualismo, demuestra la necesidad racional de las restricciones morales, que adquieren, de este modo,

un fundamento de validez universal. Al igual que el contractualismo político parte de una presunción radical contra el Estado, el poder y la comunidad política, así el contractualismo moral se caracteriza por una radical presunción en contra de la moralidad. Y aunque varias teorías normativas se han calificado como "contractualistas", ninguna de ellas se compromete con esa presunción hasta el punto en que lo hace *MA*; al contrario, usualmente pre-suponen una decisión moral, o una fuente —subjetiva u objetiva— de las restricciones morales externa (anterior) al argumento contractual.

Pero las pretensiones del contractualismo en cuanto teoría ética son limitadas. El minimalismo de las premisas garantiza la universalidad a costa de la reducción de los contenidos. El contractualismo no puede —ni lo pretende— dar cuenta de toda la riqueza moral de las personas. La moralidad justificada por el contrato se reduce a "un conjunto de disposiciones, prácticas y afectivas, que permiten a los agentes capacitados para la restricción racional, ejercitar esa capacidad comprometiéndose en empresas cooperativas encaminadas al beneficio mutuo"[3].

Amplísimas regiones compuestas por nuestros sentimientos morales no son abarcadas por el contractualismo moral; pero no por ello quedan marginadas, o reducidas a expresión más o menos irracional de unos afectos socialmente condicionados. Porque, enraizados en la disposición que nos mueve a la cooperación, y producto inmediato de la sociedad entendida como una empresa cooperativa para el beneficio mutuo, los sentimientos morales poseen también una base racional (si bien, la explicitación de la misma haría infinitamente compleja nuestra teoría).

Si las reglas morales fuesen convencionales y nuestros sentimientos morales simples expresiones subjetivas irracionales, entonces las premisas del contrato lo habrían puesto de manifiesto. La posibilidad de arribar a una moralidad que pueda entenderse como necesidad de la razón dada nuestra

---

[3] Gauthier, D., "Rational Constraints: Some Last Words", en Vallentyne, P., *Contractarianism and Rational Choice*, Cambridge, Cambridge U.P., 1991, pp. 323-330; p. 229. Aunque, como decíamos en el punto anterior, una sociedad liberal permite y fomenta una vida moral mucho más rica. en este punto estamos distinguiendo lo que la teoría puede justificar como estrictamente derivado de sus postulados, y lo que puede admitir por no ser contrario —sino coherente— con sus conclusiones.

constitución (como agentes orientados a fines y abocados a la interacción), resume el éxito (y el sentido) del contractualismo.

El contrato no es, por tanto, la historia de la construcción de la moralidad; sino la reconstrucción racional de su posibilidad, narrada como una historia. A la vez, el contrato es el mecanismo necesario para esa reconstrucción. Decir que nuestra moral es contractual significa decir que nuestros principios morales *sólo* pueden ser racionalmente justificados como producto de una negociación entre agentes auto-interesados; significa decir que nuestras reglas morales *sólo* pueden entenderse como plasmación de esos principios acordados; significa decir que el origen de nuestros sentimientos morales, si es racional, *sólo* puede pensarse en relación con la disposición a restringir el comportamiento maximizador derivada del contrato.

En definitiva, la moral por acuerdo, no es una simple convención moral. Es la expresión de una necesidad de la razón cuando se entiende como un conjunto de "racionalidades" en interacción. Y a esta luz se aprecia que la disposición moral pertenece a la esencia de un agente racional en cuanto miembro de una sociedad. De modo que podemos afirmar que el postulado individualista (que incluía el auto-interés) es superado por las conclusiones de la teoría: se muestra justamente como un simple postulado.

Pero el contractualismo no sólo refleja la necesidad racional de la cooperación y de la moralidad como disposición personal; también capta el carácter esencialmente público del ámbito moral. Porque exige metodológicamente una negociación en la que los intereses de todas y cada una las personas implicadas sean tenidos en cuenta y defendidos; de modo que el resultado de esa negociación ideal sea aceptable *ex post* por cada persona concreta. La aceptabilidad del acuerdo se basa en que éste refleja imparcialmente los intereses no-morales de cada agente en un contexto de negociación regido por los principios públicos de libertad, racionalidad individual, auto-interés y maximización; los cuales garantizan la ausencia de fraude o imparcialidad.

El resultado de la negociación, esto es, el principio de distribución justa, se constituye en principio de una razón intersubjetiva que establece el marco mínimo para la convivencia social. Ello es así en la medida en que este principio permite y regula la cooperación y es, además, criterio de moralidad

de las instituciones.

El contractualismo moral puede entenderse, en esta vertiente, como el esfuerzo por dotar de una base racional —y no meramente razonable— al conjunto mínimo de valores compartidos por los miembros de una sociedad liberal. La exploración en esta dirección nos abocaría a una reflexión filosófico-política cuyo lugar no es éste. No obstante, se nos permitirá una breve alusión a la relación entre el ideal gauthieriano de una sociedad liberal basada en el contractualismo moral, y la idea del liberalismo político apoyado en un consenso razonable en el ámbito público.

3. Ética mínima, liberalismo y el ámbito de lo público

La idea de que la ética que puede ser justificada contractualmente (o, en general, procedimentalmente) es una ética mínima ha sido blanco de incontables críticas. Ya hemos reconocido arriba que una vasta región de lo que identificamos con el "ámbito moral" no puede ser justificada directamente apelando al beneficio mutuo. Esto dota a la teoría moral liberal de cierto carácter escandaloso: ¿cómo es posible defender una doctrina ética que, ante el problema de justificar la solidaridad con los discapacitados (y en general, con todos aquellos que no pueden contribuir *nada* al esfuerzo cooperativo), simplemente calla o reconoce que su límite son los "casos normales"?[4]

---

[4] Prescindimos, obviamente, de la respuesta inmediata a este tipo de críticas, que consiste en remitir a lo que hemos dicho arriba sobre la capacidad moral de un individuo liberal. En realidad, una moral contractual sí justifica, aunque indirectamente, la solidaridad con quienes no contribuyen nada (sino que son socialmente gravosos). Una sociedad liberal es el verdadero semillero de sentimiento de comunidad humana que podría garantizar ciertos actos (individuales o sociales) rigurosamente altruistas.

Tampoco aludiremos, para no extendernos innecesariamente, al argumento de que, si se admite que el beneficio de la sociedad no es estrictamente el beneficio de la cooperación, sino que hay un "efecto multiplicador" debido a las posibilidades —inexistentes en el estado de naturaleza— que la sociedad abre (muchas de ellas en relaciones con la promoción de relaciones afectivas entre los individuos), entonces cabe pensar que la valoración de una cooperación universal sea tan grande para cada uno, que se no se considere "coste" el hecho de "incluir" en la sociedad a quienes, aparentemente, no contribuyen, pero que en realidad sí lo hacen, debido al valor que adquiere su mera presencia como miembros de la sociedad.

Creemos que las explicaciones anteriores sobre la sociedad liberal deberían ser suficientes para rechazar ese tipo de críticas. Pero supongamos que, efectivamente, una ética mínima no abarcase nada más que el ámbito del beneficio mutuo, ¿de qué escandalizarse?, ¿no es ésa, acaso, la única ética coherente con gran parte de nuestra ideología?, ¿no admitimos, explícita o implícitamente, de modo general el individualismo presente en nuestras sociedades, el relativismo y subjetivismo de los valores, el carácter predominantemente económico e instrumental de la racionalidad, la concepción contractual de las relaciones sociales?

Esos rasgos de nuestra ideología liberal tal vez sean desagradables, pero son indiscutibles: representan (o al menos están conectados con) el componente racional[5] en nuestra reflexión práctica. Aceptarlos como materia prima de las premisas del argumento moral y, a pesar de todo, construir un procedimiento que permite justificar no sólo la adopción de acuerdos justos, sino su cumplimiento, representa un esfuerzo de profundo e inesperado alcance: aunque parece que tales supuestos sólo pueden conducir —como argumenta la crítica— a una limitada "ética" del hombre económico, entendida como "maximización a largo plazo" o como "egoísmo ilustrado", sin embargo, nuestro análisis revela que ese esfuerzo supone el establecimiento de principios morales de base racional que sirven como fundamento del acuerdo mínimo compartido que mantiene una sociedad liberal.

En este nivel político, podemos conectar las conclusiones de Gauthier con las categorías del último Rawls. Porque la base racional de la moral por

---

Un argumento más contra la acusación de "inhumanidad" diría que es razonable pensar que los agentes auto-interesados diseñaran, entre las instituciones básicas de la sociedad, mecanismos de previsión para el caso de ser ellos mismos víctimas de una enfermedad o accidente. Por motivos igualmente auto-interesados es probable que cada agente quisiera prevenir también la posibilidad de no poder correr con los gastos de sus cuidados, de modo que se estableciera un sistema de ayuda mutua, como parte del compromiso originario de la cooperación (que es racional cumplir, como maximizador restringido). Una vez en marcha, este esquema de cooperación abarcaría los casos que suscitan esta crítica.

[5] Hemos de interpretar aquí el componente racional como opuesto tanto al componente moral como a lo razonable, entendido como la disposición a proponer y cumplir principios y normas que establezcan los términos imparciales de la cooperación (Cfr., respecto a esto último, Rawls, J., *Political Liberalism*, Nueva York, Columbia U.P., 1993, p. 49).

acuerdo garantiza algo que el propio Rawls requiere: que el 'consenso solapante' no sea materia de transacción o convenio[6], sino que pueda ser aceptado por todos como parte esencial (y, según los casos, "verdadera") de sus respectivas doctrinas comprensivas o concepciones de la vida. Si entendemos correctamente a Rawls en este punto, una moral contractual podría servir de modelo para ese núcleo de convicciones compartidas. El argumento contractualista, al partir de premisas racionales, explicaría por qué ese núcleo no es tanto una mera convención razonable como parte de una ética mínima racionalmente necesaria y, en esa medida, compartida por todas las personas. Por otro lado —como ya hemos apuntado— el contractualismo da razón de la disposición a pactar y cumplir acuerdos justos, es decir, explica en términos racionales el componente razonable que hay que suponer en la base de una sociedad liberal[7].

Aunque es una lectura arriesgada, que sólo proponemos como hipótesis problemática o como programa futuro de investigación, creemos que no es desencaminado conectar tan íntimamente el contractualismo moral con la reflexión política. Al fin y al cabo, el proyecto contractualista se asienta en la

---

[6] Cfr. Rawls, *Political Liberalism*, cit., p. 150 y ss. Allí Rawls demanda que el consenso compartido no sea "indiferente" o "escéptico".

[7] Cfr., sobre "lo razonable", Rawls, *op. cit.*, pp. 49 y ss. El texto de la p. 49 donde Rawls dice que "las personas son razonables en un aspecto básico cuando, entre iguales, están dispuestos a proponer principios y reglas como términos justos de la cooperación y a someterse a ellos voluntariamente, siempre que se asegure que los demás harán lo mismo", apoya —creemos— nuestra tesis sobre la función de la moral por acuerdo en una sociedad liberal, pues Gauthier muestra cómo es posible que surja, desde la pura racionalidad estratégica, una disposición condicional a la cooperación y una disposición a cumplir los acuerdos justos. Según esto, el cálculo meramente racional (justificación moral) sería anterior lógicamente a la discusión sobre términos razonables de los principios de la cooperación (justificación política). De todas formas, se trata de conexiones que sólo entrevemos problemáticamente, y que merecen un desarrollo posterior más minucioso.

Por otro lado, nuestra lectura del contrato moral liberal no sólo modifica o complementa el liberalismo político de Rawls, sino que también podría emplearse como complemento de la teoría política basada en el racionalismo crítico de Popper (tal como queda recogido por Jiménez Perona en *Entre el liberalismo y la socialdemocracia*, Barcelona, Anthropos, 1993,p. 211). Popper —dice Jiménez Perona— no tiene ningún recato en afirmar que su racionalismo "se basa en una fe irracional en la actitud de razonabilidad" (Cfr. Popper, K., *el desarrollo del conocimiento científico. Conjeturas y refutaciones*, Buenos Aires, Paidos, 1979, p. 411). Nuestra opinión es que el contractualismo moral, al fundamentar racionalmente la razonabilidad, añade a las teorías políticas liberales, tanto de Rawls como de Popper, el *plus* de racionalidad que demandan, pero del que carecen.

conciencia de que los asuntos prácticos no son asuntos que pueda resolver el individuo aislado, sino que tienen un carácter esencialmente público. Dicho carácter es captado, en el nivel político, por la idea de un consenso razonable; en el nivel de los principios morales, por la idea de un acuerdo racional ideal; y en el nivel de la motivación individual, por la posibilidad de justificar las propias acciones ante los demás sobre términos que ellos no pudieran razonablemente rechazar (es decir, que podrían haber sido producto de un acuerdo unánime)[8].

4. El contrato moral como ideología: crítica final

Al final de *MA* Gauthier retorna a la inquietante pregunta sobre el alcance de su proyecto moral: ¿No se reducirá éste a una reafirmación, más sofisticada, del egoísmo, travestido en aceptación "sincera" de las restricciones morales? El fundamento instrumental de la ética liberal, ¿es compatible con un proyecto moral? ¿No habremos justificado una moralidad que resulta ser una ficción, una impostura?

Esta duda surge del excesivo énfasis puesto en *MA* sobre la idea de que la moralidad es un *artificio* o *estratagema* para superar el dilema de la racionalidad[9]. Como sabemos, al inicio del desarrollo de su teoría Gauthier no confió especialmente en el alcance de una moralidad basada en la racionalidad estratégica. Sólo en la parte final de *MA*, aparece la idea de que la cooperación no posee un simple valor instrumental, sino que adquirirá valor intrínseco para

[8] Sobre éste último nivel de la motivación individual, Cfr. Scanlon, "Contractualism and Utilitarianism", en Sen y Williams, *Utilitarianism and Beyond*, Cambridge, Cambridge U.P., 1982, pp. 103-128; y Rawls, *Political Liberalism*, cit., pp. 49-50, nota 2.

[9] En un trabajo posterior sobre este problema —"Value, Reasons and the Sense of Justice" (en Frey (ed.), *Value, Welfare and Morality*, Nueva York, Cambridge U.P., 1993, pp. 180-20; p. 180)— Gauthier escribe: "También se podría considerar la moralidad no como un artificio, sino como un conjunto de disposiciones naturales. Aunque introduje tal idea en *MA*, no tenía papel alguno en los argumentos centrales de mi libro."

un individuo liberal[10]. Queda allí patente que los elementos esenciales de una moral contractual —el reconocimiento de los derechos de los otros y el compromiso de someterse a los principios acordados— no son en absoluto simulacros y, "si son requeridos por la interacción racional, entonces, partiendo de la concepción del comportamiento racional más estrecha y moralmente neutra que pueda plausiblemente obtenerse, sabremos desarrollar una ética que no diverja excesivamente de nuestras convicciones morales más asentadas y cuyo núcleo continúe siendo la inviolabilidad y la dignidad de la personalidad moral"[11].

Sin embargo, sería ingenuo confiar en el poder de convicción del contractualismo moral. El escéptico no quedará persuadido de que se trate de algo más que de una hábil añagaza del egoísmo; mientras el "moralista" no admitirá, por principio, que las restricciones al auto-interés puedan surgir de la racionalidad estratégica (y si lo admitiera, negaría que tales restricciones merecieran la dignidad de "morales"). Ante este previsible inconformismo, Gauthier insiste lacónicamente en el sentido de su teoría:

> "*MA* es un desafío a la profética sentencia de Nietzsche, *'a medida que la voluntad de verdad gane así auto-conciencia ... la moralidad irá pereciendo poco a poco'*. Es un intento de escribir

---

[10] Idea que es desarrollada especialmente en la sección 9 (pp. 197-199) de "Value, Reasons and the Sense of Justice", cit.

En relación a la misma idea, considérese el siguiente párrafo de Scanlon ("Contractualism and Utilitarianism", cit., p. 128): "A veces se dice que la moralidad es una estratagema para protegernos unos de otros. Según el contractualismo, esta idea es en parte verdad, pero incompleta en un sentido importante. Nuestra inquietud por proteger nuestros intereses principales tendrá un importante efecto sobre lo que podríamos acordar razonablemente. Así, tendrá un importante efecto sobre el contenido de la moralidad si el contractualismo es correcto. En la medida en que esta moralidad sea observada, aquellos intereses serán promovidos. Si no tuviéramos ningún deseo de justificar nuestras acciones ante otros sobre una base que razonablemente aceptable por ellos, la esperanza de ganar esta protección nos daría razones para tratar de infundir este deseo en otros, quizás a través de una hipnosis o condicionamiento en masa, incluso aunque ello significase que nosotros también lo adquiriríamos. Pero dado que ya contamos con este deseo, nuestro interés en la moralidad es menos instrumental."

[11] Gauthier, D., "Between Hobbes and Rawls", en Gauthier y Sugden (eds.), *Rationality, Justice and the Social Contract*, Ann Arbor, University of Michigan Press, 1993, pp. 24-39; p. 39.

> una teoría moral para adultos, para personas conscientes de vivir en un mundo post-antropomórfico, post-teocéntrico, post-tecnocrático. Es un intento de disipar el miedo, o la sospecha, o la esperanza, de que sin un fundamento en el valor objetivo o en la razón objetiva, en la simpatía o la sociabilidad, la empresa moral está destinada al fracaso."[12]

La moral contractual pretende mostrar el éxito de la 'empresa moral' desde las únicas premisas que puede aceptar el individuo liberal. Tal vez se trate de una victoria pírrica —así lo entenderán muchos—, pero defendemos que se trata de la única respuesta compatible con los supuestos ideológicos de la sociedad liberal y, por tanto, la única teoría que puede esperar servir como base de justificación para el comportamiento moral del individuo liberal.

La moral contractual está, por tanto —y pese al universalismo de sus conclusiones—, conectada con la sociedad liberal. No pretende situarse fuera del marco histórico en que ha nacido: su validez se circunscribe a un ámbito en que la racionalidad, la individualidad, el orden social y la acción práctica se conciben de una cierta forma. Si estas concepciones variasen, posiblemente también debería hacerlo el paradigma de justificación moral de la acción.

Gauthier sostiene que el contrato moral es una contribución —quizá destinada a ser superada— en el camino de secularización, autonomía y libertad que inició la ilustración. Esto tiene una doble lectura: por un lado expresa la convicción de que el contractualismo moral es un proyecto inacabado, que requiere modificaciones tal vez constantes; por otro, supone la admisión explícita de que se trata de una ética para sociedades modernas, democráticas, liberales y de mercado.

Ahora bien, reconocer que la ética representada por el contractualismo está históricamente situada, no significa negar que responda a la necesidad de un fundamento intersubjetivamente válido de la acción moral. Y al aportar ese fundamento, los principios morales son reconocidos por cada individuo como necesarios, como principios normativos universales de la interacción.

---

[12] "Moral Artifice", *Canadian Journal of Philosophy*, vol. 18, nº 2, pp. 385-418; p. 385.

Hasta qué punto la teoría tiene éxito es difícil de valorar. A lo largo de nuestro trabajo, hemos mostrado las articulaciones más quebradizas de la misma; muchas de ellas ya superadas o modificadas por el propio autor. Respecto de los argumentos concretos; juzgar la validez de todos ellos excede con mucho nuestra competencia. La mayoría nos parecen, sin embargo, plausibles. Sobre la teoría en conjunto, compartimos la opinión de Buchanan, quien la pondera mucho como propuesta moral —como argumento en defensa de que las personas deben (por razones morales tanto como prudenciales) adoptar la postura moral incorporada en los principios de la ética contractual[13]. En cuanto al valor de *MA* como fundamentación de un orden social liberal, ya hemos expuesto nuestra opinión arriba.

Para nosotros, el intento de David Gauthier permite una esperanza, porque muestra que el contractualismo moral es, como proyecto, factible; que el pluralismo inherente a la sociedad liberal no destituye la autoridad de la razón en materia moral. Porque la razón sigue expresando sus demandas universales, ahora a través del mecanismo heurístico de una negociación ideal sobre los principios de justicia; una negociación que recoge el carácter esencialmente público y originariamente prudencial de la racionalidad práctica.

El contractualismo nos enseña que, como individuos liberales, debemos revisar nuestra cultura moral; debemos —asegura Fishkin— aprender a esperar menos de una teoría moral[14]. Si nos aplicamos a ello, tal vez encontremos en el argumento contractualista el fundamento intersubjetivo de la ética que puede permitir la convivencia en cualquier sociedad plural entendida como una unión de personas libres y autónomas —una ética mínima, sí; pero universal.

Madrid, noviembre de 1995

---

[13] Cfr. Buchanan, J., "The Gauthier Enterprise", en Paul, E.F. *et al.* (eds.), *The New Social Contract*, Oxford, Blackwell, 1988, pp. 75-93; p. 75.

[14] Cfr. Fishkin, J., "Liberal Theory and the Problem of Justification", en Pennock,J.R. y Chapman, J.W., *Justification*, Nueva York, New York U.P., 1986, pp. 207-231; p. 227.

Apéndice

Apéndice

Breve revista de la recepción de la teoría de Gauthier en la filosofía española

No es necesario recalcar la influencia que tuvo, en los años setenta, la *Teoría de la Justicia*, de John Rawls. Esta obra revolucionó la Filosofía Política y provocó una conmoción en los campos afines, como la Ética, la Teoría del Estado, etc. En todo el mundo, los especialistas interesados en enfoques analíticos de la Filosofía Política, en estudios utilitaristas, en la Filosofía Jurídica norteamericana, etc., se sintieron atraídos (o inquietados) por la teoría y el método de Rawls. Por otro lado, el progreso de la Teoría de la Decisión Racional se extendió por todo el mundo gracias a la Ciencia Económica que, a través de la Escuela del *'Public Choice'* (Elección Pública) amplió y profundizó notablemente el campo de la Economía Política, hasta convertirla en una de las principales teorías normativas sobre el comportamiento humano.

La revitalización de estos campos en los Estados Unidos hizo que muchos científicos y filósofos de todos los continentes prestasen una renovada atención a la tradición filosófica anglosajona (radicada principalmente en América del Norte). Los trabajos de Harsanyi, Buchanan, Nozick, Parfit, Elster y otros teóricos fueron (y aún son) discutidos y estudiados con enorme interés.

Morals by Agreement apareció en el marco de esta atmósfera de interés por la filosofía política y moral anglosajona basada en los avances de la Teoría de la Decisión Racional. Una atmósfera presente también en la filosofía española y latinoamericana.

No es extraño que la primera recepción de *MA* en España e Hispanoamérica tuviera lugar entre quienes cultivaban las dos líneas de pensamiento que definen la obra: la Filosofía Política y la Teoría de la Decisión Racional. Eran, en su mayoría, filósofos del Derecho, algunos economistas y *pocos* filósofos morales (los interesados en aspectos interdisciplinares de la Teoría de la Decisión Racional relacionados con el utilitarismo, el neo-contractualismo, los

problemas de identidad personal, etc.).

a) Primeras lecturas

Hasta donde yo conozco, *MA* fue discutida por primera vez de dos modos bien diferentes: en la Universidad de Alicante (en su Departamento de Filosofía del Derecho), el profesor Manuel Atienza organizó un debate sobre *MA*, cuyas ponencias fueron publicadas en el número seis de la revista *Doxa*. Hacia la misma época, en la Universidad Complutense de Madrid (en la Facultad de Filosofía) la Srta. Blanca Rodríguez preparaba su tesis doctoral bajo la dirección del catedrático de Ética, profesor Gilberto Gutiérrez. La tesis se tituló *Moralidad y cooperación racional* y estaba influida por *MA* (aunque Rodríguez disentía de Gauthier en aspectos fundamentales de su teoría). Dedicaré una breve comentario a estas primeras recepciones de *MA*, que representan paradigmáticamente las lecturas ius-filosóficas y éticas del contractualismo moral liberal. También incluiré en este epígrafe dos artículos posteriores (de 1991 y 1992) porque están formulados como "primeras reacciones" ante *MA*, y un artículo más que ejemplifica un "uso técnico" del modelo de negociación presentado en *MA* como solución posible para juegos cooperativos.

La tesis doctoral de Rodríguez fue defendida en la primavera de 1990. Ella no empleó literatura secundaria sobre *MA*, de modo que su lectura puede ser considerada "de primera mano". Se puede decir que Rodríguez "usa" conceptos e ideas de *MA* para su propio argumento. En mi opinión, comprende adecuadamente algunos puntos difíciles, como el papel de la salvaguardia lockeana en la determinación de la posición inicial de negociación, o el análisis de las condiciones para el cumplimiento del pacto. Por otro lado, Rodríguez estudia en detalle las diferencias entre los modelos de negociación racional desarrollados por Zeuthen-Nash, Harsanyi y Gauthier. Concluye que, a pesar de la precisión matemática de los modelos previos, la teoría de Gauthier resulta intuitivamente más aceptable. No obstante, el estudio de *MA* en la tesis de Rodríguez se desarrolla para defender un argumento utilitarista (semejante al de Harsanyi), y esto le lleva a ciertos malentendidos. Toma *MA* como un ensayo sobre una posible solución para el problema práctico (individual) de la cooperación racional, pero pierde de vista la dimensión contractualista de la obra (esto es, ella probablemente no vio la *necesidad* de la cooperación misma,

ni la relación que la teoría establece entre un principio de la negociación racional e imparcial y la justicia). Desde el punto de vista que Rodríguez adoptó, *MA* resultaba una propuesta interesante, pero completamente ajena a la ética. Pensó que el capítulo final (la discusión sobre el individuo liberal como una persona que valora la cooperación por sí misma más que, instrumentalmente, por su utilidad para alcanzar fines egoístas) era un intento de dar razones para cumplir el pacto y, lógicamente, no confió en un mecanismo de cooperación racional cuyo mantenimiento demandaba lo que ella interpretó como un "cambio en los valores" de cierto número de individuos (un número mínimo de personas *debían* convertirse en agentes cooperativos convencidos y sinceros independientemente de lo que los demás hicieran). Esta comprensión del último capítulo (y de la obra completa, en cuanto se entiende supeditada al mismo) trivializa el alcance de *MA*.

La importancia del estudio de Rodríguez reside en que se desarrolló desde dentro de una discusión ética, en relación con problemas prácticos concretos. Paradójicamente, las lecturas más influyentes de *MA* se dieron en un campo un tanto extraño a la obra misma: la Filosofía del Derecho. Como he comentado, los autores neo-contractualistas eran mucho mejor conocidos entre algunos filósofos del Derecho y la Política, cuyo interés en el pensamiento contractualista les llevó a iniciar la discusión pública de *MA*, como una nueva contribución a esa tradición. El debate sostenido en la Universidad de Alicante trajo consigo la publicación de artículos de Martín D. Farrell, Ruth Zimmerling y Albert Calsamiglia (además de la traducción de "¿Por qué contractualismo?" de Gauthier)[1]. Todos ellos presentaron lecturas indudablemente interesantes y reflexivas, pero, desde mi punto de vista, un tanto superficiales.

Por ejemplo, el artículo de Martín D. Farrell ("El dilema de Gauthier") plantea un supuesto dilema en estos términos: o se establecen restricciones racionales desde un punto de partida no-moral (pero entonces se tratará de restricciones prudenciales), o se establecen restricciones morales (pero irracionales en el sentido anterior), partiendo de una concepción justificatoria de la racionalidad que debe incluir ya criterios morales. Este pretendido dilema está tomado, más o menos, del análisis de Joseph Mendola ("Gauthier's *Morals by Agreement* and Two Kinds of Rationality", en *Ethics*, 97 [julio 1987], pp.

[1] Estos artículos constituyeron la parte monográfica del número 6 (1989) de la revista *DOXA*, al que nos referíamos arriba.

765-774)[2], que no deja de ser un tanto superficial. No negamos que el dilema pueda existir,pero los argumentos ofrecidos por Farrell son tan débiles (si es que pueden ser considerados argumentos) que permiten una fácil superación mediante una comprensión algo más profunda de *MA*.

Otro ejemplo es el artículo de Ruth Zimmerling ("La pregunta del tonto y la respuesta de Gauthier", pp. 49-76). Zimmerling presenta la obra de Gauthier de un modo brillante, como respuesta a la objeción del *Tonto*. Además, ofrece una descripción muy precisa de la teoría y detecta, con un agudo análisis, sus puntos débiles. Pero cuando trata de desplegar su crítica, se enreda en una maraña de argumentos generales sobre la posibilidad y carácter de la justificación racional en ética, y el tipo de justificación que Gauthier supuestamente defiende.

De los tres artículos publicados en *Doxa*, el de Calsamiglia es probablemente el más lúcido (sin negar el valor de los otros dos). Se titula "Un egoísta colectivo: ensayo sobre el individualismo según Gauthier" (pp. 77-94). Realiza también una clarificadora descripción del objetivo de *MA* y expone ciertos comentarios críticos. Calsamiglia reconoce el interés del intento de fundar una moralidad sobre la sola base del auto-interés, pero critica lo que denomina "el mordisco normativo", representado, entre otros, por el postulado de la translucidez (al justificar la racionalidad de la disposición a convertirse en un maximizador restringido), la idea de la *igual* racionalidad y, por supuesto, el uso de la salvaguardia lockeana como límite de la interacción natural. De acuerdo con Calsamiglia, este "mordisco normativo", que provendría de una fuente distinta del simple auto-interés, va alejando progresivamente la teoría de su punto de partida, y reduciría su plausibilidad. Calsamiglia sostiene que, a pesar de los esfuerzos de Gauthier, la moralidad permanece alienada del sujeto, porque sus fuentes se sitúan —fuera del auto-interés individual— en "elementos substanciales de lo que se denomina la dignidad o la autonomía de la persona", donde deberíamos buscar el manantial del "mordisco normativo" que va diluyendo el objetivo inicial de la teoría.

Se puede decir que el debate mantenido en la universidad de Alicante abrió la discusión pública sobre *MA* como obra representante de un enfoque contractualista liberal de la moralidad. Después de 1989 es posible encontrar referencias a *MA* en diversos trabajos relacionados con la ética, teoría política, Teoría de la Decisión, etc. Pero antes de comentar algunas de estas referencias,

---

[2] El mismo dilema constituye la tesis principal de la obra de Jung Soon Park *Contractarian Liberal Ethics and the Theory of Rational Choice* (Nueva York, Peter Lang, 1992), uno de los análisis más profundos sobre el contractualismo de Gauthier.

debemos decir algo sobre tres textos que pueden incluirse entre las "primeras lecturas" de *MA*.

El primero de ellos es un trabajo estrictamente académico de M. Pilar González Altable, leído como ponencia en el I *Congreso Iberoamericano de Estudios Utilitaristas*, en septiembre de 1991. Fue publicado posteriormente en *Telos* (vol. I, nº 2, junio 1992, pp. 111-125) con el siguiente título: "El contractualismo liberal de David Gauthier. Contractualismo vs. utilitarismo". El artículo es una buena exposición de *MA* —quizá excesivamente dependiente de los artículos publicados en *Doxa*, especialmente el de Calsamiglia. Reproduce el argumento de Gauthier contra el utilitarismo, tomándolo principalmente de *MA* y de "¿Por qué contractualismo?". La conclusión valora positivamente el componente liberal de la teoría contractualista de Gauthier, pero sigue a J.S. Fishkin y R. Hardin al señalar que incluso aunque la teoría de Gauthier logre justificar una concepción de la justicia distributiva, no consigue justificar su concepción de "justicia esencial".

El segundo texto al que me gustaría referirme es una breve recensión, escrita por José Montoya, de la Universidad de Valencia, ("D. Gauthier o Hobbes sin Leviatán", *Revista de Filosofía*, 3ª época, vol. IV (1995), nº 5, pp. 199-205). Se trata probablemente de uno de los más perspicaces comentarios sobre *MA*. El profesor Montoya sitúa la teoría de Gauthier en la tradición hobbesiana, como un intento de resolver uno de los problemas clásicos de la filosofía moral moderna, a saber, la reconciliación del auto-interés con el comportamiento correcto. Este problema surge únicamente *en el marco* del radical individualismo moderno (que implicó la destrucción de la idea del hombre como un animal social). En mi opinión, Montoya ofrece una explicación esclarecedora de algunas de las etapas más importantes del argumento de Gauthier: la idea del mercado perfectamente competitivo como zona exenta de moralidad, el surgimiento de la cooperación, la determinación de un principio distributivo, etc. Y todo su análisis se despliega en relación con dos paradigmas del pensamiento moral: el paradigma humeano y el hobbesiano. Finalmente Montoya valora algo que ningún otro crítico había resaltado: el hecho de que Gauthier no trata de establecer una teoría ética *sub specie aeternitatis,* sino mostrar un importante aspecto de nuestra ideología moderna. En este sentido, Montoya cree que Gauthier es consciente del valor histórico de su contribución. Esta lectura es coherente con algunos de los comentarios y dudas que aparecen al final de *MA*, y permite una muy interesante —aunque cabría preguntarse hasta qué punto fiel— interpretación de la teoría.

Para finalizar esta sección sobre las primeras lecturas, creo que es interesante resaltar la recepción de la teoría de Gauthier entre los teóricos de

juegos hispanohablantes. Un ejemplo de ésta se encuentra en el mismo número de *Doxa* en que se publicó el debate sobre *MA*. Me refiero al artículo de Julia Barragán "Las reglas de la cooperación" (*Doxa*, 6, 1989, pp. 329-384). Se trata de un largo estudio sobre posibles soluciones para los juegos cooperativos. La profesora Barragán analiza diversas propuestas teóricas y, entre ellas, la idea de la maximización restringida. Considera que la cooperación basada en la maximización restringida sería inestable porque dependería de la exclusión coactiva de los maximizadores directos y de la improbable condición de la "translucidez". No obstante, Barragán acepta el análisis gauthieriano de las condiciones de la cooperación y ensaya con aplicaciones de la teoría de la negociación racional como solución a algunos problemas políticos. En suma, es crítica con el resultado de la obra de Gauthier como solución de los problemas de cooperación, pero toma en consideración su contribución a las teorías de juegos y de la negociación racional. Este solo hecho es importante, sobre todo si se tiene en cuenta la estrecha colaboración entre la profesora Barragán y J.C. Harsanyi en el campo de la acción colectiva[3].

b) Reflexiones y críticas posteriores

Tras las primeras lecturas y discusiones, *MA* se convirtió en una reconocida contribución a la ética contemporánea, y comenzó a incluirse en muchos de los trabajos relacionados con éticas contractualistas o procedimentalistas. La filosofía moral española está dominada por la ética discursiva y, en buena parte, por la "filosofía continental" —aunque la influencia de la tradición anglosajona no es, como ya se ha dicho, despreciable. Este hecho puede ser la causa de la visión crítica de *MA* que predomina en nuestra literatura. Sin embargo, es importante señalar que la obra de Gauthier es tomada frecuentemente como una piedra de toque, un punto de referencia polémico liberal en medio de un océano de teorías constructivistas de inspiración kantiana (Habermas, Apel, Rawls).

En este clima de general aceptación y/o discusión, se pueden encontrar referencias más o menos extensas a *MA* en muchos libros y artículos especializados, de modo que es imposible ofrecer una relación exhaustiva. Nos centraremos sólo en algunas opiniones presentadas por teóricos reconocidos en

[3] Una colaboración plasmada, por citar sólo un ejemplo reciente, en Griffin, Barragán, Harsanyi y Bardón, *Ética y política en la decisión pública*, Caracas, Angria, 1993.

libros recientes. Como excepción, consideraremos con cierto detenimiento la obra de J.C. Bayón Mohíno, por su extraordinaria densidad.

Expondremos primero el comentario que encontramos en *Ética constructiva y autonomía personal* (Madrid, Tecnos, 1992), de J. Rubio Carracedo. Sostiene Rubio Carracedo que existen dos paradigmas en la ética contemporánea: el basado en una concepción estratégica de la racionalidad y el basado en una concepción comunicativa. El primer paradigma está representad por el primer Rawls, Baier, Grice y, eminentemente, por Gauthier. El paradigma comunicativo corresponde al sustrato de la concepción ética de Habermas, Apel y, en un sentido algo diferente, del último Rawls. Rubio Carracedo explica, siguiendo a Apel, que una concepción estratégica de la racionalidad nunca podrá servir como base para la ética porque obstaculiza, en vez de facilitar, un significativo y verdadero consenso basado en una comunicación sincera. La posibilidad de una acción no estratégica —una posibilidad que el mismo hecho de la comunicación evidencia— despejaría el camino hacia un fundamento más profundo de la ética. Frente a la ética comunicativa, el paradigma estratégico le parece a Rubio Carracedo (y a sus aliados intelectuales) muy limitado.

Desde un punto de vista diferente, Martín D. Farrell intenta criticar el principio de concesión relativa *minimax* arguyendo que representa una mala alternativa al segundo principio rawlsiano de la justicia. Farrell incluye esta crítica en su libro *La filosofía del liberalismo* (Madrid, Centro de Estudios Constitucionales, 1992). Su objeción no se dirige tanto al fundamento teórico del principio —que considera "atractivo"— cuanto a la plausibilidad de su aplicación práctica. En la práctica resulta ser —en opinión de Farrell— un pobre sustituto del principio de la diferencia. Obviamente, la conclusión de Farrell es que las carencias del principio rawlsiano pueden ser superadas mediante su desarrollo como principio representativo de la línea dominante de la filosofía liberal; mientras, el principio de Gauthier no pasa, en su consideración, de una crítica —fallida, por cierto— al de Rawls.

Otro profesor latinoamericano, Carlos S. Nino, fue autor del artículo "Ética analítica en la actualidad", incluido en *Concepciones de la ética* (Madrid, Trotta, 1992). La contribución de Nino forma parte de una obra que pretende ser algo así como un tratado general sobre la ética contemporánea. Y es interesante notar que, una vez más, la teoría de Gauthier se trata desde el punto de vista de su relación con la de Rawls. en este caso, *MA* se contempla como uno de los posibles modos de desarrollar e interpretar la *Teoría de la justicia*. Nino afirma que *MA* puede considerarse una teoría (metaética) sobre

la naturaleza de los principios morales: éstos pueden ser concebidos como teoremas de la Teoría de la Decisión Racional, deducidos del axioma de la maximización. En este sentido, *MA* representaría una de las posibles formas de interpretar la obra de Rawls.

Como hemos dicho, estos ejemplos sirven como representantes del modo habitual en que la teoría de Gauthier es interpretada y discutida. Pero si hemos de seleccionar un tratamiento más pormenorizado de *MA*, hemos de volvernos hacia J.C. Bayón Mohíno, cuya tesis doctoral (publicada como *La normatividad del derecho: deber jurídico y razones para la acción*, Madrid, Centro de Estudios Constitucionales, 1991) dedica un buen número de páginas al análisis de *MA*.

Debemos decir que no compartimos algunas de las ideas de fondo defendidas por Bayón Mohíno, ni sus conclusiones. Pero se trata de un estudio extraordinariamente completo, y merece especial atención.

Bayón Mohíno dedica la primera parte de su tesis al estudio de las razones para la acción. Distingue tres niveles de razones: deseos (o preferencias), intereses y razones morales. *MA* sirve para ejemplificar cómo los intereses (razones de segundo orden) pueden ampliarse hasta comprender también a las razones morales, superando, al mismo tiempo, el fracaso del auto-interés como razón para actuar. Desde este punto de vista, *MA* queda situada en un difícil equilibrio entre las razones internas (deseos, prudencia) y las razones externas y objetivas (deberes y reglas morales).

Escribe Bayón Mohíno que, de acuerdo con Gauthier, la moralidad puede definirse como "el conjunto de restricciones a la satisfacción del propio interés, adoptado por razones de autointerés, que permitiría alcanzar resultados Pareto-óptimos y cuya aceptación haría posible al mismo tiempo a cada individuo obtener una satisfacción de sus intereses más completa que si actuara siempre como un maximizador irrestricto" (p. 161). Este texto —punto de partida de su crítica— es fiel sólo a una parte de la concepción gauthieriana de la moral. Se aprecia cierto sesgo en la lectura que Bayón Mohíno ofrecerá, por que no menciona conceptos tales como "cooperación", "negociación", "conformidad", etc.; da por supuesto que la concepción moral de Gauthier es puramente instrumental, lo cual está expresamente negado en el último capítulo de *MA*.

Si éste el punto de partida; seguidamente Bayón Mohíno considera que la teoría de Gauthier es susceptible de dos tipos de críticas: una interna y otra externa. Las objeciones internas aceptarían las premisas para cuestionar el argumento mismo: negarían la posibilidad de deducir restricciones al auto-

interés a partir de razones auto-interesadas. La debilidad del argumento respecto de este tipo de crítica se localizaría en los tres aspectos siguientes: el proceso de negociación, la posibilidad de aceptación del mismo o conformidad, y el papel de la salvaguardia lockeana (*proviso*). En el proceso de negociación Bayón Mohíno detecta una ilegítima idea de igualdad. Tal idea podría haberle parecido legítima si hubiera rastreado su origen —que se halla en el postulado de igual racionalidad. Pero, al compartimentar el análisis, considera que la imparcialidad *en* el proceso de negociación representa una premisa moral injustificada[4].

Sus objeciones sobre el cumplimiento del contrato y la salvaguardia nos parecen más sólidas (aunque tal vez descansen en una interpretación cuestionable de la teoría como un todo). Para ahorrarnos detalles[5], digamos que su conclusión es que sería económicamente racional estar dispuesto a aceptar y cumplir algunos acuerdos incluso aunque no estuvieran basados en una posición original imparcial (restringida por la salvaguardia). Si esto es verdad, entonces se pondría en duda la conexión entre el resultado de una negociación racional y la moral —ya que esta conexión depende de la *necesidad* racional (económica, diría Bayón Mohíno) del principio de Concesión Relativa *Minimax*— y, por otro lado, no habría ninguna razón determinante para cumplir unos contratos en vez de otros. Gauthier habría justificado la racionalidad (en el sentido de "prudencia") de cumplir (ciertos) pactos, en la medida en que pueden ser beneficiosos, pero no habría demostrado que cierto acuerdo ideal pudiera servir como fuente de un criterio moral.

Aún va más allá Bayón Mohíno en su intento de criticar la idea misma que subyace al cumplimiento del contrato: cree que "adoptar una disposición" (lo que exige la maximización restringida) tiene implicaciones y dificultades que Gauthier no reconoce. Emplea argumentos de Bernard Williams (*Ethics and the Limits of Philosophy*, Londres, Fontana/Collins, 1985) y E.F. McClennen ("Constrained Maximization and Resolute Choice", *Social Philosophy and Policy*, 5, pp. 95-118, reimpreso en E.F. Paul *et al.* (eds.), *The New Social Contract*, misma paginación) para cuestionar la posibilidad de una transformación como la requerida por la adopción de una nueva disposición.

La conclusión de esta primera parte de su análisis es que Gauthier no

[4] También puede entenderse la crítica de Bayón Mohino como una reivindicación del concepto (aceptado desde Nash) de "capacidad negociadora de las partes". Pero el modelo de Gauthier no olvida esta variable, lo que ocurre es que la despeja al dar por sentado que, entre agentes perfecta e igualmente racionales, todos poseen la misma "habilidad negociadora".

[5] El argumento se encuentra en *La normatividad del derecho...*, pp. 171 y ss.

prueba la racionalidad de cumplir los acuerdos pactados, así que no puede probar la racionalidad de concertarlos. Mucho menos de lograr un acuerdo de las características requeridas por la teoría moral contractual. En la opinión de Bayón Mohíno, Gauthier no ha logrado mostrar que la moralidad pueda emerger como una restricción racional a partir de las premisas no-morales de la decisión racional[6].

Esta debe considerarse la conclusión de las "críticas internas". Si nos centramos en la "externa", hallaremos que Bayón Mohíno ofrece un argumento que puede calificarse como original (aunque desencaminado, desde nuestro punto de vista). En vez de argumentar que la concepción de la moralidad como una restricción al auto-interés es pobre y contradice nuestras ideas comunes sobre los sentimientos morales; en vez de argumentar que la concepción instrumental de la racionalidad no capta la riqueza de nuestro razonamiento práctico; en vez de tratar de postular una concepción de la moralidad basada en un concepto substantivo (histórico, kantiano, discursivo o cualquier otro) de la racionalidad; o en vez de proclamar que *MA* funda una moralidad estrecha que no incluiría la mayoría de lo que tradicionalmente se consideran deberes morales; en vez de cualquiera de estas líneas argumentativas, Bayón Mohíno afirma que el concepto de moralidad de Gauthier es auto-contradictorio.

Esta tesis está basada en la idea de que, si la teoría de Gauthier fuera correcta, entonces cualquier agente tendría una razón para aceptar los límites (a la persecución individual del auto-interés) impuestos por la moralidad contractual, y *ninguna* razón para aceptar cualesquiera otras restricciones. De esta forma, cualquier creencia moral distinta de la apoyada en el principio de concesión relativa *minimax* y en la salvaguardia lockeana, sería irracional. Lo cual es un modo de decir que, para cualquier individuo, existen *razones* morales para actuar correctas e incorrectas: correctas si están apoyadas por la teoría de Gauthier, incorrectas en otro caso. En opinión de Bayón Mohíno, esta última aserción contradice la teoría del valor defendida en el capítulo II de *MA*. Y esta contradicción desacreditaría la obra como un todo.

Bayón Mohíno no entiende por qué, si el valor es relativo al agente, no

[6] Sería enormemente prolijo analizar los fundamentos de la crítica de Bayón Mohíno. Creemos que ese análisis revelaría que, si bien gran parte de sus conclusiones son acertadas, algunos de los razonamientos en que las funda carecen de solidez, por estar apoyados en simples malentendidos. Por poner un ejemplo, su crítica a la conformidad estricta parte de la base de que el modelo de racionalidad de Gauthier *se identifica* con la racionalidad económica. Para nosotros quedó claro, en el capítulo II, 2, que el modelo de racionalidad de Gauthier tiene que ver con la racionalidad económica, pero no se puede identificar con ella.

puede considerarse *cualquier* preferencia como una buena razón para actuar (incluidas las preferencias consistentes en no maximizar el auto-interés, o en no adherirse al principio *minimax*). ¿Por qué —pregunta— habría que calificar una preferencia como *irracional* simplemente a causa de que difiere de la "preferencias morales" aprobadas por la teoría? Obviamente, Bayón Mohíno lee *MA* como una defensa de un criterio axiológico concreto; un criterio según el cual valorar la verdadera racionalidad (y moralidad) de las acciones. Y niega que tal criterio pueda ser deducido de premisas tan contradictorias con él (la concepción relativista del valor y la concepción económica —instrumental, diríamos nosotros— de la racionalidad). Desde su punto de vista, en la conclusión de *MA* está implicado un tipo de racionalidad no sólo diferente, sino además incompatible con la racionalidad económica aceptada como premisa; y no percibe el puente deductivo entre ambas.

En nuestra opinión, el estudio resumido señala de modo insuperable las dificultades principales de *MA*. Se hace eco de las críticas mejor construidas y trata de añadir pregnantes razonamientos dirigidos contra el proyecto de *MA*. Sin embargo, creemos que Bayón Mohíno no aprehende adecuadamente el argumento de la obra (tal vez subordina el examen de *MA* a la defensa de sus propias tesis). Su análisis convierte a *MA* en un conjunto de ideas desarticuladas[7], evidentemente criticables. Los razonamientos de sus críticas —inspirados en las revisiones de *MA* que siguieron inmediatamente a la publicación de la obra— nos parecen unos mejor, y otros peor fundados. No trataremos todas sus posibles debilidades[8]; simplemente diremos algo sobre las que creemos son las razones generales de su perspectiva crítica.

La primera razón es su ya mencionada comprensión parcial de la obra, que creemos se muestra en su definición de moralidad. Bayón Mohíno trata de leer *MA* como un modo de evitar la auto-frustración (o auto-refutación) colectiva de la concepción maximizadora de la racionalidad. Desde este punto de vista, considera la maximización restringida y los principios de la moralidad simplemente como el "verdadero interés" de cada individuo, que le permitirían alcanzar resultados colectivos óptimos (pero contingentes). No capta ninguna

[7] Por cierto que esto es lo que el mismo Gauthier pensaba de su obra no mucho antes de publicarla; así, en el prefacio nos dice que "de hecho, en un tiempo pensé publicar gran parte del presente libro como un estudio de un conjunto de interconexiones conceptuales sin pretender que globalmente constituyera la teoría moral correcta." (*MA*, pp. V-VI).

[8] Máxime teniendo en cuenta que muchas de ellas se han tratado ya en el cuerpo de la Tesis, bien refiriéndonos directamente a Bayón Mohíno, bien al replicar a otros críticos.

dimensión moral en esos mecanismos. La misma localización del análisis de *MA* en su libro revela este punto de vista: en vez de situar *MA* entre los ejemplos de las razones para la acción de tercer orden (razones morales), lo sitúa como una posibilidad de superar las insuficiencias de las razones de segundo orden (intereses).

Se puede decir que el punto de vista adoptado por Bayón Mohíno no se atiene al objetivo expreso de *MA*. Además, si examinamos su contribución en detalle, observamos que mezcla en sus críticas diversos niveles argumentales. Ello es especialmente claro cuando toma las premisas y conclusiones de *MA* como si estuvieran en el mismo plano. Trata igualmente el nivel de las decisiones (y razones para la acción) individuales y el nivel de la discusión normativa de los principios[9]. Y, por otro lado, olvida que la conclusión, esto es, la racionalidad como maximización restringida, *incluye* a la maximización directa como parte suya, de modo que no puede ser contradictoria con ella.

No creemos que estos malentendidos reduzcan el valor de las críticas desarrolladas por Bayón Mohíno; tan sólo pretendemos que expliquen lo que nos parece un enfoque injustificadamente hipercrítico.

c) Conclusión

De todo lo dicho, se podría deducir que la obra de Gauthier ha sido comentada en España exclusivamente por sus detractores. Quizá no hemos enfatizado suficientemente la admiración general hacia el esfuerzo intelectual que representa; un elogio tanto más destacable cuanto proviene de teóricos no precisamente afines a la ideología de Gauthier.

A pesar de esto, es imposible negar que la mayoría de las reacciones hacia *MA* han sido escépticas y críticas. Quizá las expectativas de teóricos provenientes de campos como la Decisión Pública o la Filosofía del Derecho no eran adecuadas a lo que la obra podía y pretendía ofrecer. Algunos otros tal vez esperaban "otra *Teoría de la Justicia*", es decir, una teoría política, y ni siquiera captaron el alcance de la obra como teoría moral.

---

[9] Nos parece clave distinguir el nivel de las premisas, donde se sitúa la concepción subjetiva del valor y el de las conclusiones normativas, donde se sitúan las restricciones morales derivadas de la negociación y el pacto. Esta distinción elimina el cortocircuito que Bayón Mohíno cree detectar en la obra. Las preferencias (y valores) individuales y los principios intersubjetivos no sólo no son contradictorios, sino que el argumento de *MA* tiende a hacer ver con claridad por qué son esencialmente compatibles.

Nuestra conclusión es que, entre las diferentes tradiciones y enfoques de los autores que han reflexionado sobre *MA*, no se ha encontrado hasta ahora ningún intento de comprensión hondamente filosófico. Cuando éste ha sido bosquejado (como en el caso de Montoya), ha resultado interesante e iluminador. Nosotros hemos pretendido seguir ese camino. Nuestro trabajo ha pretendido y pretende únicamente resaltar las profundas implicaciones filosóficas del proyecto de *MA*. Incluso aunque la teoría contractualista de la moralidad no tenga éxito en la forma intentada en *MA*, merece ser desarrollada (quizá por otros caminos) porque es, en nuestra opinión, una de las (pocas) vías filosóficas para revitalizar el ideal moderno de una moralidad universal y racional, frente (o tal vez en paralelo) a las tesis del neo-aristotelismo o el llamado, en general, comunitarismo.

Agra Romero, María José, "Etica neo-contractualista", en Victoria Camps (ed.), *Concepciones de la ética*, Madrid, Trotta, 1992, pp. 247-268.

Aguiar, Fernando (comp.), *Intereses individuales y acción colectiva*, Madrid, ed. Pablo Iglesias, 1991.

Aguiar, Fernando, "La lógica de la cooperación", en Aguiar (comp.) *Intereses individuales y acción colectiva*, Madrid, Pablo Iglesias, 1991.

Apel, Karl-Otto, *La transformación de la filosofía (I Análisis del lenguaje, semiótica y hermenéutica) (II, El a priori de la comunidad de comunicación)*, Madrid, Taurus, 1985. (2 vols.)

Aroso Linhares, José Manuel, "Habermas y la argumentación jurídica", en *Revista de la Facultad de Derecho de la Universidad Complutense*, nº 79, curso 1991/92, p. 27-53.

Ashcraft, Richard (ed.), *John Locke. Critical Assessments. (Vol. III, Politics)*, Londres, Routledge, 1991.

Axelrod, R., *La evolución de la cooperación*, Madrid, Alianza, 1987.

Baier, Annette C., "Pilgrim's Progress", *Canadian Journal of Philosophy*, Vol. 18, nº 2, Junio 1988, pp. 315-330.

Baier, Kurt, "Justification in Ethics", en J.R. Pennock y J.W. Chapman (eds.), *Justification* (Nomos XXVIII), Nueva York, Nueva York U. P., 1986, pp. 3-27.

Baier, Kurt, *The Moral Point of View: A Rational Basis of Ethics*, Ithaca (N.Y.), Cornell University Press, 1958.

Barragán, Julia, "Las reglas de la cooperación", *DOXA* 6 (1989), pp. 329-384.

Barragán, Julia, "El poder normativo de las autoexcepciones", *Relea*, nº 0, abril 1995, pp. 24-41.

Barry, Brian, *Theories of Justice*, Londres, Harvester, 1989.

Bayón Mohino, Juan Carlos, *La normatividad del derecho: deber jurídico y razones para la acción*, Madrid, Centro de Estudios Constitucionales, 1991.

Béjar Merino, Helena, *El ámbito íntimo: privacidad, individualismo y modernidad*, Madrid, Alianza, 1988.

Benhabib, Seyla, "El otro generalizado y el otro concreto: La controversia Kohlberg-Gilligan y la teoría feminista", en *Teoría feminista y teoría crítica*, Valencia, Ediciones Alfons el Magnanim, 1990, pp.119-149.

Betegón, Jerónimo, y Juan Ramón de Páramo (dir. y coords.), *Derecho y moral. Ensayos analíticos*, Barcelona, Ariel, 1990

Bhargava, Rajeev, *Individualism in Social Science. Forms and Limits of a Methodology*, Oxford, Clarendon, 1992.

Bobbio, Norberto, *Thomas Hobbes*, Barcelona, Plaza y Janés, 1991.

Boucher, D. y Kelly, P. (eds.), *The Social Contract from Hobbes to Rawls*, Londres y Nueva York, Routledge, 1994.

Brandt, Richard B., *Ethical Theory*, Englewood Cliffs (N.J.), Prentice Hall, 1959. Versión española: *Teoría ética*, Madrid, Alianza, 1982 (trad. de Esperanza Guisán).

Brandt, Richard, "The Concept of Rationality in Ethical and Political Theory", en Pennock, J.R. y Chapman, J.W. (eds.), *Human Nature in Politics*, Nueva York, New York U.P., 1977, pp. 256-279.

Brandt, Richard B., *A Theory of the Good and the Right*, Oxford, Clarendon, 1979.

Braybrooke, David, "Social Contract Theory's Fanciest Flight", en *Ethics*, 97, 4, Julio 1987, pp. 750-764.

Buchanan, Allen, "Justice and Reciprocity versus Subject-Centered Justice", en *Philosophy and Public Affairs*, 19 (1990), pp. 227-257.

Buchanan, James M. y Tullock, Gordon, *The Calculus of Consent. Logical Foundations of Constitutional Democracy*, Ann Arbor, University of Michigan Press, 1965.

Buchanan, James M., *The Limits of Liberty. Between Anarchy and Leviathan*, Chicago, University of Chicago Press, 1975.

Buchanan, James M., "The Gauthier Enterprise", en Paul, E.F. *et al.* (eds.), *The New Social Contract*, Oxford, Blackwell, 1988, pp. 75-94.

Calsamiglia, Albert, "Un egoista colectivo. Ensayo sobre el individualismo según Gauthier", en *DOXA*, 6 (1989), pp. 77-94.

Campbell, Richmond y Sowden, Lanning (eds.), *Paradoxes of Rationality and Cooperation: Prisoner's Dilemma and Newcomb's Problem*, The University of British Columbia Press, 1985.

Campbell, Richmond, "Moral Justification and Freedom", en *The Journal of Philosophy*, 85 (1988), pp. 192-213.

Camps, Victoria, *Ética, Retórica, Política*, Madrid, Alianza, 1988

Camps, Victoria, Guariglia, O. y Salmerón, F. (eds.), *Concepciones de la ética*, Madrid, Trotta, 1992.

Castro Leiva, Luis (compilador), *El liberalismo como problema*, Caracas, Monte Avila Latinoamericana, 1992.

Clayton, P. y Knapp, S., "Ethics and Rationality", en *American Philosophical Quarterly*, vol. 30, nº 2, abril 1993.

Cortina, Adela, *Ética mínima*, Madrid, Tecnos, 1986.

Cortina, Adela, *Ética sin moral*, Madrid, Tecnos, 1992 (2ª ed.).

Danielson, Peter, "The Visible Hand of Morality", en *Canadian Journal of Philosophy*, vol. 18, nº 2, Junio 1988, pp. 357-384.

Davis, Morton D., *Introducción a la Teoría de juegos*, Madrid, Alianza, 1986.

De Jassay, Anthony, *Social Contract, Free Ride: A Study of the Public Goods Problem*, Oxford, Oxford University Press, 1989.

De Jasay, Anthony, *Choice, Contract, Consent: A Restatement of Liberalism*, Londres, Institute of Economic Affaires, 1991.

De Jasay, Anthony, *El estado. La lógica del poder político*, Madrid, Alianza, 1993 (trad. Rafael Caparrós Valderrama).

De-Shalit, Avner, "Bargaining with the Not-Yet-Born? Gauthier's Contractarian Theory of Inter-generational Justice and its Limitations", en *International Journal of Moral and Social Studies*, otoño 1990, pp. 221-234.

Diaz, Elias, *Ética contra política. Los intelectuales y el poder*, Madrid, Centro de Estudios Constitucionales, 1990.

Dunn, John, *La teoría política de occidente ante el futuro*, Mexico, Fondo de Cultura Económica, 1981.

Dunn, John, *The Political Thought of John Locke*, Cambridge, Cambridge University Press, 1986.

Dunn, John, *Rethinking Modern Political Theory*, Cambridge, Cambridge University Press, 1986.

Dworkin, Ronald, "Liberalism" en Hampshire, S. (ed.), *Public and Private Morality*, Cambridge, Cambridge University Press, 1978.

Dyke, C., "The Vices of Altruism", en *Ethics*, 81 (1971) p. 241-251.

Elster, Jon, *Uvas amargas*, Barcelona, Península, 1988 (trad. E. Lynch).

Elster, Jon, *Ulises y las sirenas*, Mexico, Fondo de Cultura Económica, 1989 (trad. Juan José Utrilla).

Elster, Jon, "Racionalidad, moralidad y acción colectiva", en Aguiar (comp.), *Intereses individuales y acción colectiva*, Madrid, Pablo Iglesias, 1991.

Farrell, Martín D., "El dilema de Gauthier", en *DOXA*, 6 (1989), pp. 39-48.

Farrell, Martín Diego, *La filosofía del liberalismo*, Madrid, Centro de Estudios Constitucionales, 1992.

Fishkin, James S., "Liberal Theory and the Problem of Justification", en J.R. Pennock y J.W. Chapman (eds.) *Justification* (*Nomos XXVIII*), Nueva York, New York U.P., 1986, pp. 207-231.

Franssen, Maarten, "Constrained Maximization Reconsidered - An Elaboration and Critique of Gauthier's Modelling of Rational Cooperation in a Single Prisoner's Dilemma", en *Synthese*, vol. 101, nº 2, Noviembre 1994, pp. 249-272.

Fraser, N. y Gordon, L., "Contrato *versus* caridad", en *Isegoría*, nº 6, 1992, p. 65-82 (trad. Pedro Francés Gómez).

Gauthier, David P., *Practical Reasoning. The Structure and Foundations of Prudential and Moral Arguments and their Exemplification in Discourse*, Oxford, Clarendon, 1963.

———, "Rule-utilitarianism and Randomization", en *Analysis*, vol. 25, nº 3, Enero de 1965, pp. 168-69.

———, "How Decisions Are Caused", en *The Journal of Philosophy*, vol. LXIV, nº 5, Marzo, 1967.

———, "Moore's Naturalistic Fallacy", *American Philosophical Quarterly*, vol. 4, nº 4, octubre 1967, pp. 315-320.

———, "Morality and Advantage", en *Philosophical Review*, 76 (1967), pp. 460-475. Reimpreso en Gauthier (ed.), *Morality and Rational Self-Interest*, pp. 166-180. Versión española "La moral y la ventaja", en Joseph Raz (comp.), *Razonamiento práctico*, México, F.C.E., 1986 (trad. de Juan José Utrilla).

———, "Progress and Happiness: A Utilitarian Reconsideration", *Ethics*, 78 (1967), pp. 77-82.

———, "The Unity of Wisdom and Temperance", *Journal of the History of Philosophy*, 6 (1968), pp. 157-159.

———, "Hare's Debtors", en *Mind*, 77 (1968), pp. 400-405.

———, "How Decisions are Caused (But not Predicted)", en *The Journal of Philosophy*, vol 65, nº 6, Marzo 1968, pp. 170-1.

———, *The Logic of Leviathan. The Moral and Political Theory of Thomas Hobbes*, Oxford, Clarendon Press, 1969 (primera reimpresión 1979).

———, "Yet Another Hobbes", en *Inquiry*, 12 (1969), pp. 449-473.

——— (ed.), *Morality and Rational Self-Interest*, Englewood Cliffs (Nueva Jersey), Prentice Hall, 1970.

———, "The Impossibility of Rational Egoism", en *The Journal of Philosophy*, vol. LXXI, nº 14, Agosto de 1974, pp. 439-456.

———, "Rational Cooperation", en *Noûs*, 8 (1974), pp. 53-65.

———, "Justice and Natural Endowment: Toward a Critique of Rawls' Ideological Framework", en *Social Theory and Practice*, 3 (1974), pp. 3-26. Reimpreso en Gauthier, *Moral Dealing*, pp. 150-170.

———, "Reason and Maximization", en *Canadian Journal of Philosophy*, 4 (1975), pp. 411-433. Reimpreso en Gauthier, *Moral Dealing*, pp. 209-234.

———, "Coordination", en *Dialogue*, 14 (1975), pp. 195-221. Reimpreso en Gauthier, *Moral Dealing*, pp. 274-297.

———, "Why Ought One Obey God? Reflections on Hobbes and Locke", *Canadian Journal of Philosophy*, 7 (1977), pp. 425-446. Reimpreso en Gauthier, *Moral Dealing*, pp. 24-44.

———, "The Social Contract as Ideology", en *Philosophy and Public Affairs*, 6 (Otoño 1977), pp. 130-164.

———, "The Social Contract: Individual Decision or Collective Bargain?", en Hooker, C.A., Leach, J.J. y McClennen, E.F. (eds.), *Foundations and Applications of Decision Theory*, Dorsrecht, Reidel, 1978, pp. 690-706.

———, "Social Choice and Distributive Justice", en *Philosophia* n° 7, 1978, pp. 239-153.

———, "Economic Rationality and Moral Constraints", *Midwest Studies in Philosophy*, 3 (1978), pp. 75-96.

———, "Bargaining Our Way IntoMorality: A Do-It-Yourself Primer", *Philosophical Exchange*, 2 (1979), pp. 14-27.

———, "David Hume: Contractarian", *Philosophical Review*, 88 (1979), pp. 3-38. Reimpreso en Gauthier, *Moral Dealing*, pp. 45-76.

———, "The Politics of Redemption", *Revue de l'Université d'Ottawa*, 49 (1979), pp. 329-356. Reimpreso en Gauthier, *Moral Dealing*, pp. 77-109.

———, "Thomas Hobbes: Moral Theorist", en *Journal of Philosophy*, 76 (1979), pp. 541-559. Reimpreso en Gauthier, *Moral Dealing*, pp. 11-23.

———, "The Irrationality of Choosing Egoism -A Reply to Eshelman", en *Canadian Journal of Philosophy*, vol. X, n° 2, junio 1980, pp. 179-187.

———, "Justified Inequality?", *Dialogue*, 21 (1982), pp. 431-443.

———, "No Need for Morality: The Case of the Competitive Market", *Philosophical Exchange*, 3 (1982), pp. 41-54.

———, "Three against Justice: The Foole, the Sensible Knave and the Lydian Shepherd", *Midwest Studies in Philosophy*, 7 (1982), pp. 11-29. Reimpeso en Gauthier, *Moral Dealing*, pp. 129-149.

———, "Deterrence, Maximization and Rationality", en *Ethics*, 94 (1984), pp. 474-495. Reimpreso en Gauthier, *Moral Dealing*, pp. 298-324.

———, "Bargaining and Justice", *Social Philosophy and Policy*, 2 (1985), pp. 29-47. Reimpreso en Gauthier, *Moral Dealing*, pp. 187-204.

———, "The Unity of Reason: A Subversive Reinterpretation of Kant", en *Ethics*, 96 (1985), pp. 74-88. Reimpreso en Gauthier, *Moral Dealing*, pp. 110-127.

———, "The Incompleat Egoist", en McMurrin, S.M. (ed.), *The Tanner Lectures on Human Values*, vol. 5, Salt Lake, University of Utah Press, 1985, pp. 67-119. Reimpreso en Gauthier, *Moral Dealing*, pp. 234-273.

———, "Maximization Constrained: The Rationality of Cooperation", en Richmond, C. y Lanning, S. (eds.), *Paradoxes of Rationality and Cooperation*, Vancouver, University of British Columbia Press, 1985, pp. 75-94.

———, *Morals by Agreement*, Oxford, Clarendon, 1986. Reimpreso en 1987 y 1988. Versión española *La moral por acuerdo*, Barcelona, Gedisa, 1994 (trad. Alcira Bixio).

———, "Reason to be Moral?", *Synthese*, 72 (1987), pp. 5-27.

———, "Taming Leviathan", en *Philosophy and Public Affairs*, 16 (1987), pp. 280-298.

———, "Reply to Wolfram", en *Philosophical Books*, vol. 28, nº 3, julio 1987, pp. 134-139.

———, "Moral Artifice", *Canadian Journal of Philosophy*, vol. 18, nº 2 (1988), pp. 385-418.

———, "Hobbes's Social Contract", en Rogers G.A.J. y Ryan, Alan (eds.), *Perspectives on Thomas Hobbes*, Oxford, Clarendon, 1988, pp. 125-152.

———, "Morality, Rational Choice, and Semantic Representation: A Reply to My Critics", en Paul, E.F., *et al.* (eds.), *The New Social Contract*, Oxford, Blackwell, 1988, pp. 173-221.

———, "Why Contractarianism?", Cambridge U.P., Nueva York, 1991. Versión española: "¿Por qué contractualismo?", *DOXA*, 6 (1989), pp. 19-38, (trad Silvia Mendlewicz y A. Calsamiglia).

———, *Moral Dealing. Contract, Ethics and Reason*, Ithaca, Cornell University Press, 1990.

———, "Economic Man and the Rational Reasoner", en James H. Nichols, Jr. y Colin Wright (eds.), *From Political Economy to Economics - And Back?*, Los Angeles, Institute for Contemporary Studies Press, 1990.

———, "Artificial Virtues and the Sensible Knave", en Tweyman, S. (ed.), *David Hume. Critical Assessments*, Londres, Routledge, 1990, vol. VI, pp. 129-154.

———, "Constituting Democracy", en D. Copp, J. Hampton y J.E. Roemer (eds.), *The Idea of Democracy*, Nueva York, Cambridge University Press, 1993, pp. 314-334.

———, "Value, reasons, and the sense of justice", en Frey, R.G. (ed.) *Value, Welfare and Morality*, Nueva York, Cambridge U.P., 1993, pp. 180-208.

——— y Sugden, R., *Rationality, Justice and the Social Contract. Themes from Morals by Agreement*, Ann Arbor, University of Michigan Press, 1993.

———, "Constituting Democracy", en Copp, D., Hampton, J. y Roemer, J.E. (eds.), *The Idea of Democracy*, Nueva York, Cambridge University Press, 1993, pp. 314-334.

———, "Assure and Threaten", *Ethics*, 104 (Julio 1994) pp. 690-721.

———, "Breaking Up: An Essay on Secession", *Canadian Journal of Philosophy*, vol 24, nº 3, Sept. 1994, pp. 357-372.

———, "Public Reason", en *Social Philosophy and Policy*, 12: 1 (invierno 1995), pp. 19-42.

Gewirth, Alan, "La base y el contenido de los derchos humanos", en Betegón, J. y Páramo J. R. (coords), *Derecho y moral. Ensayos analíticos*, Barcelona, Ariel, 1990, pp. 125-145.

Glasgow, W. D., "Ethical Egoism Again", en *Ethics*, 82 (1971) pp. 65-71.

Glucksmann, André, *El undecimo mandamiento ¿Es posible ser moral?*, Barcelona, Peninsula, 1993.

González Altable, Mª Pilar, "El contractuallismo liberal de D. Gauthier. Contractualismo vs. utilitarismo", en *Telos*, vol I, nº 2, Junio 1992 pp. 111-125.

Gough, J.W., *The Social Contract A critical Study of its Development*, Oxford, Clarendon, 1957, segunda edición (1ª de 1936).

Granovetter, Mark, "Modelos de umbral de conducta colectiva", en Aguiar, *Intereses individuales y acción colectiva*, Madrid, Pablo Iglesias, 1991.

Gray, John, *Liberalismo*, Madrid, Alianza, 1994 (trad. María Teresa de Mucha)

Grice, G.R., *The Grounds of Moral Judgment*, Cambridge, Cambridge University Press, 1967.

Grocio, Hugo, *De Iure Belli ac Pacis*, Madrid, Centro de Estudios Constitucionales, 1987 (ed. de P. Merino Gómez).

Gutiérrez López, Gilberto, "Racionalidad consecuencialista y restricciones deontológicas", en Javier Muguerza *et al.*, *El fundamento de los derechos humanos*, Madrid, Debate,

1989, pp. 195-202.

Habermas, Jurgen, "¿Cómo es posible la legitimidad por vía de legalidad?", en *DOXA*, 5 (1988), pp. 21-45.

Habermas, Jurgen, *Teoría de la acción comunicativa. Complementos y estudios previos*, Madrid, Cátedra, 1989.

Habermas, Jürgen, *Conciencia moral y acción comunicativa*, Barcelona, Península, 1985, (trad. Ramón García Cotarelo).

Hampshire, Stuart (comp.), *Moral pública y privada*, México, Fondo de Cultura Económica, 1983.

Hampton, Jean, *Hobbes and The Social Contract Tradition*, Cambridge, Cambridge U.P., 1986.

Hampton, Jean, "The Moral Commitments of Liberalism", en D. Copp, J. Hampton y J.E. Roemer (eds.), *The Idea of Democracy*, Nueva York, Cambridge University Press, 1993, pp. 292-313.

Hampton, Jean, "Can we Agree on Morals", en *Canadian Journal of Philosophy*, vol 18, nº 2, Junio 1988, pp. 331-356.

Hare, R.M., *Freedom and Reason*, Oxford, Oxford U.P., 1963.

Hare, R.M., *El lenguaje de la moral*, México, U.N.A.M., 1975 (trad. de Genaro R. Garrido y Eduardo A. Rabossi).

Harman, Gilbert, *The Nature of Morality: An Introduction to Ethics*, Nueva York, Oxford University Press, 1977.

Harsanyi, John C., *Essays on Ethics, Social Behavior, and Scientific Explanation*, Dordrecht, Reidel Publishing Company, 1976.

Harsanyi, John C. "Morality and the Theory of Rational Behaviour", en A. Sen y B. Williams (eds.), *Utilitarianism and Beyond*, Nueva York, Cambridge U.P., 1982 pp. 39-62.

Hartogh, Govert den, "The Rationality of Conditional Cooperation", en *Erkenntnis*, 38 (1993), pp. 405-427.

Hausman, Daniel M., "Are Markets Morally Free Zones?", *Philosophy and Public Affairs*, vol. 18, nº 4 (otoño 1989), pp. 317-333.

Hayek, Friedrich A., *Derecho, Legislación y Libertad*, Madrid, Unión Editorial,

1978-1982 (trad. L. Reig Albiol).

Hayek, Friedrich A., *La fatal arrogancia*, Madrid, Unión Editorial, 1990 (trad. L. Reig Albiol).

Hayek, Friedrich A., *Nuevos estudios en filosofía, política, economía e historia de las ideas*, Buenos Aires, Eudeba, 1991 (trad. M.I. Albes).

Hobbes, Thomas, *Leviathan (English Works, vol. 3)*, Londres, John Bohn, 1966 (ed. de Sir William Molesworth).

Hobbes, Thomas, *El Ciudadano*, Madrid, Debate/CSIC, 1993 (ed. bilingüe de Joaquín Rodríguez Feo).

Hope, V.M., *Virtue by Consensus. The moral Philosophy of Hutcheson Hume and Adam Smith*, Oxford, Clarendon, 1989.

Hume, David, *A Treatise of Human Nature*, Oxford, Oxford University Press, 1978 (2ª ed. de L.A. Selby-Bigge, rev. por P.H. Nidditch).

Jimenez Perona, Angeles, *Entre el liberalismo y la socialdemocracia. Popper y la sociedad abierta*, Barcelona Anthropos, 1993.

Kalai, E. y Smorodonsky, M., "Other Solutions to Nash's Bargaining Problem", *Econometrica*, 43 (1975), pp. 513-518.

Kant, Immanuel, "En torno al tópico: 'tal vez eso sea correcto en teoría, pero no sirve para la práctica'" , en Kant, *Teoría y Práctica*, Madrid, Tecnos, 1986, (trad. M.Francisco Pérez López y Roberto Rodríguez Aramayo).

Kant, Immanuel, *La paz perpetua*, Madrid, Tecnos, 1985 (trad. Joaquín Abellán).

Kant, Immanuel, *Primera introducción a la Crítica del Juicio*, Madrid, Visor, 1987 (trad. José Luis Zalabardo).

Kant, Immanuel, *Crítica de la Razón Práctica*, Madrid, Librería General de Victoriano Suárez, 1913 (trad. de E. Miñana y Villagrasa y M. García Morente).

Kant, Immanuel, *Fundamentación de la metafísica de las costumbres*, Madrid, Espasa, 1990 (9ª ed. de Luis Martínez de Velasco).

Kavka, Gregory, *Hobbesian Moral and Political Theory*, Princeton (N.J.), Princeton University Press, 1986.

Kern, L. y Müller H.P., *La justicia: ¿Discurso o mercado?. los nuevos enfoques de la teoría contractualista*, Barcelona, Gedisa, 1992.

Klosko, George, *The Development of Plato's Political Theory*, Nueva York y Londres, Methuen, 1986.

Kohlberg, L., Levine, C. y Hewer, A., *Moral Stages: A Current Formulation and a Response to Critics*, Basilea y Nueva York, Karger, 1983.

Kolm, Serge-Christophe, *Le contrat social libéral. Philosophie et practique du libéralisme*, Paris, Presses Universitaires de France, 1985.

Koons, Robert C., "Gauthier and the Rationality of Justice", en *Philosophical Studies*, vol. 76, nº 1, octubre, 1994, pp. 1-26.

Kraus, Jody S., y Coleman, Jules L., "Morality and the Theory of Rational Choice", *Ethics* 97 (Julio 1987), pp. 715-749.

Kraus, Jody S., *The Limits of Hobbesian Contractarianism*, Nueva York, Cambridge University Press, 1993.

Laurent, Alan, *Histoire de l'individualisme*, Paris, Presses Universitaires de France, 1993.

Locke, John, *Essays on the Law of Nature*, Oxford, Clarendon, 1988 (ed. por W: von Leyden).

Locke, John, *Segundo tratado sobre el gobierno civil*, Madrid, Alianza Editorial, 1990 (trad. Carlos Mellizo).

Lottenbach, Hans, "Expected Utility and Constrined Maximization: Problems of Compatibility", en *Erkenntnis*, 41 (1994), pp. 37-48.

Lucash, F.S. y Shklar, J.N., *Justice and Equality Here and Now*, Ithaca, Cornell U.P., 1986.

Luce, D.R. y Raiffa, H., *Games and Decisions*, Nueva York, John Wiley & Sons, 1957.

Lycos, Kimon, *Plato on Justice and Power. Reading Book I of Plato's Republic*, Londres,Mac Millan, 1987.

MacIntyre, Alasdair, *Tras la virtud*, Barcelona, Crítica, 1987 (trad. Amelia Valcarcel).

Macpherson, C.B., *The political Theory of Possessive Individualism. Hobbes to Locke*, Oxford, Oxford University Press, 1967.

Mariana, Juan de, *La dignidad real y la educación del rey (De Rege et Regis Institutione)*, Madrid, Centro de Estudios Constitucionales, 1981 (ed. de Luis Sánchez Agesta).

Marrone, Pierpaolo, "Contrato e utilità: il caso di Hume", en *Filosofia Politica*, Bolonia, vol. VI, nº 2, 1992.

Martínez Marzoa, Felipe, *Desconocida raíz común (estudio sobre la teoría kantiana de lo bello)*, Madrid, Visor, 1987.

McClennen, Edward F., *Rationality and Dynamic Choice. Foundational Explorations*, Cambridge, Cambridge University Press, 1990.

McLean, Iain, *Public Choice*, Oxford, Blackwell, 1987.

Mendola, Joseph, "Gauthier's Morals by Agreement and Two kinds of Rationality", en *Ethics*, 97, 4, Julio 1987, pp. 765-774.

Messerly, John, "The essence of David Gauthier's Moral Philosophy", en *Kinesis*, vol. 18, nº 2, invierno 1992, pp. 39-55.

Mill, John Stuart, *On the Logic of the Moral Sciences. A System of Logic, Book VI*, New York, Bobbs-Merrill, 1965 (ed. de Henry M. Magid).

Molina, Luis de, *La teoría del justo precio*, Madrid, Editora Nacional, 1981 (ed. de Francisco Gómez Camacho).

Moore, G.E., *Ensayos éticos*, Barcelona, Paidos, 1993 (trad. Carme Castells Auleda).

Moore, Margaret, "Gauthier's Contractarian Morality", en Boucher, D. y Kelly, P. (eds.), *The Social Contract from Hobbes to Rawls*, Londres y Nueva York, Routedge, 1994, pp. 211-225.

Morris, Brian, *Western Conceptions of the Individual*, Nueva York, Berg, 1991.

Muguerza, Javier, "La alternativa del disenso", en Javier Muguerza *et al.*, *El fundamento de los derechos humanos*, Ed. Debate, Madrid, 1989, pp. 19-56.

Muguerza, Javier, *Desde la perplejidad (Ensayos sobre la ética, la razón y el diálogo)*, México, Fondo de Cultura Económica, 1990.

Nagel, Thomas, *The Possibility of Altruism*, Princeton (New Jersey), Princeton University Press, 1978 (originalmente en Oxford, Clarendon, 1970).

Nash, John F., "The Bargaining Problem", *Econometrica*, 18 (1950), pp. 155-162.

Nino, Carlos S., "Constructivismo epistemológico: entre Rawls y Habermas", *Doxa*, 5 (1988), pp. 87-105.

Nino, Carlos S., "Etica analítica en la actualidad", en Victoria Camps *et al.* (eds.) *Concepciones de la ética*, Madrid, Trotta, 1992, pp. 131-152.

Nozick, Robert, *Anarquía, Estado y Utopía*, México, Fondo de Cultura Económica, 1988 (trad. Rolando Tamayo).

Olson, Robert G., *The Morality of Self-Interest*, Nueva York, Harcourt Brace World, 1965.

Oppenheim, Felix E., "Justification in Ethics: its Limitations", en J.R. Pennock y J.W. Chapman (eds.), *Justification* (Nomos XXVIII), Nueva York, New York U. P., 1986, pp. 28-32.

Parfit, Derek, *Reasons and Persons*, Oxford, Oxford University Press, 1986.

Parfit, Derek, "Prudencia, Moralidad y el dilema del prisionero" en *Diálogo Filosofico*, nº 13, 1989, pp. 4-30.

Park, Jung Soon, *Contractatrian Liberal Ethics and the Theory of Rational Choice*, Nueva York, Peter Lang, 1992.

Pateman, Carole, *The Sexual Contract*, Stanford (California), Stanford University Press, 1988.

Paul, E.F. *et al.* (eds.), *The New Social Contract. Essays on Gauthier*, Oxford, Blackwell, 1988.

Platón, *La República*, Madrid, Instituto de Estudios Políticos, 1970 (ed. bilingüe en tres vols. de J. M. Pabon y M. Fernández Galiano).

Popper, Karl R., *La miseria del historicismo*, Madrid, Taurus, 1961 (trad. Pedro Schwartz)

Rawls, John, *Teoría de la Justicia*, México, Fondo de Cultura Económica, primera reimpresión 1985 (primera edición 1979).

Rawls, John, "Social Unity and Primary Goods", en A. Sen y B. Williams (eds.), *Utilitarianism and Beyond*, Nueva York, Cambridge U.P., 1982, pp. 159-185.

Rawls, John, *Political Liberalism*, Nueva York, Columbia University Press, 1993.

Resnik, Michael D., *Choices. An Introduction to Decision Theory*, Minneeapolis, University of Minnesota Press, 1987.

Richmond, Campbell y Lanning, Sowden (eds.), *Paradoxes of Rationality and Coopera-*

tion. Prisoner's dilemma and Newcomb's Problem, Vancouver, University of British Columbia Press, 1985.

Riley, Patrick, *Will and Political Legitimacy. A critical exposition of Social Contract Theory in Hobbes, Locke, Rousseau, Kant and Hegel*, Cambridge (Mass.), Harvard University Press, 1982.

Ripstein, Arthur, "Gauthier's Liberal Individual", en *Dialogue*, 28 (1989), pp. 63-76.

Rodilla, Migel Angel, "Buchanan, Nozick, Rawls: Variaciones sobre el estado de naturaleza", en *Anuario de Filosofía del Derecho*, 2 (1985), pp. 229-284.

Rogers, G.A.J. y Ryan, Alan (eds.), *Perspectives on Thomas Hobbes*, Oxford, Clarendon, 1988.

Rousseau, Jean-Jacques, *El contrato social o principios de derecho político*, Madrid, Tecnos, 1988 (trad. María José Villaverde).

Rubio Carracedo, José, *Paradigmas de la política. Del estado justo al estado legítimo (Platón, Marx, Rawls, Nozick)*, Barcelona, Anthropos, 1990.

Rubio Carracedo, José, *Ética constructiva y autonomía personal*, Madrid, Tecnos, 1992.

Rubio Carracedo, José, "Los dos paradigmas de la ética: estrategia y comunicación", en José Rubio Carracedo, *Ética constructiva y autonomía personal*, Madrid, Tecnos, 1992, pp. 59-85.

Ryan, Alan, "Hobbes and Individualism", en Rogers, G.A.J. y Ryan, Alan, *Perspectives on Thomas Hobbes*, Oxford, Clarendon, 1988, pp. 81-105.

Sandel, Michael (ed.), *Liberalism and its Critics*, Nueva York, New York University Press, 1992 (2ª reimpresión; 1ª ed. de 1984).

Sauvé, Kevin, "Gauthier, Property Rights, and Future Generations", *Canadian Journal of Philosophy*, vol. 25, nº 2, junio, 1995, pp. 163-176.

Scanlon, T.M., "Contractualism and Utilitarianism", en Sen y Williams (eds.), *Utilitarianism and Beyond*, Nueva York, Cambridge University Press, 1982, pp. 103-128.

Sen, Amartya, *Sobre ética y economía*, Madrid, Alianza, 1989 (trad. Angeles Conde).

Shklar, Judith N., "Injustice, Injury, and Inequality: An Introduction", en Lucash, Frank S. (ed.) *Justice and Equality Here and Now*, Ithaca, Cornell University Press, 1986, pp 13-34.

Spinoza, Baruch de, *Tratado teológico político*, Madrid, Alianza, 1986 (ed. de Atilano Dominguez).

Thiebaut, Carlos, *Los límites de la comunidad*, Madrid, Centro de Estudios Constitucionales, 1992.

Torek, Paul, "liberties, Not Rights: Gauthier and Nozick on Property", en *Social Theory and Practice*, vol. 20, nº 3 (Otoño 1994).

Toulmin, Stephen E., *El puesto de la razón en la ética*, Madrid, Alianza, 1979 (trad. I. F. Ariza).

Vallentyne, Peter (ed.), *Contractarianism and Rational Choice. Essays on David Gauthier's 'Morals by Agreement'*, Nueva York, Cambridge University Press, 1991.

Vallespín Oña, Fernando, *Nuevas Teorías del Contrato social: John Rawls, Robert Nozick, James Buchanan*, Madrid, Alianza, 1985.

Van Parijs, Philippe, *¿Qué es una sociedad justa?*, Barcelona, Ariel, 1993, (Trad. Juana A. Bignozzi).

Varoufakis, Yanis, "Modern and Postmodern Challenges to Game Theory", en *Erkenntnis*, 38 (1993), pp. 371-404.

Versenyi, Laszlo, "Is Ethical Egoism Really Inconsistent?", en *Ethics*, 80 (1970), p. 240-242.

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

Watkins, John, "Second Thoughts on Self-Interest and Morality", en Campbell, R y sowden, L. (eds.), *Paradoxes of Rationality and Cooperation*, Vancouver, University of British Columbia Press, 1985, pp. 59-74.

Windolph, F. Lyman, *Leviathan and Natural Law*, Princeton, Princeton University Press, 1951.

Wolfram, Sybil, "Morals by Agreement", en *Philosophical Books*, vol. 28, nº 3, Julio de 1987, pp. 129-134.

Zimmerling, Ruth, "La pregunta del tonto y la respuesta de Gauthier", en *DOXA*, 6 (1989), pp. 49-76.