

## LA PREGUNTA DEL TONTO Y LA RESPUESTA DE GAUTHIER<sup>1</sup>

*The Foole hath sayd in his heart, there is no such thing as Justice; and sometimes also with his tongue; seriously alleaging, that every mans conservation, and contentment, being committed to his own care, there could be no reason, why every man might not do what he thought conduced thereunto: and therefore also to make, or not make; keep, or not keep Covenants, was not against Reason, when it conduced to ones benefit. He does not therein deny, that there be Covenants; and that they are sometimes broken, sometimes kept; and that such breach of them may be called Injustice, and the observance of them Justice: but he questioneth, whether Injustice, taking away ghe feare of God, (for the same, Foole hath said in his heart there is no God,) may not sometimes stand with that Reason, which dictateth to every man his own good... and if it be not against Reason, it is not against Justice: or else Justice is not to be approved for good.*

Hobbes, 1651, cap. 15

### I. INTRODUCCIÓN

**D**e vez en cuando, hasta un Tonto tiene un momento lúcido y se le ocurre, si no una respuesta, al menos una pregunta que le asegura un lugar en el panteón de la historia de las ideas normalmente reservado a los Genios. Esto es lo que ha pasado, por ejemplo, con el Tonto de Hobbes, que desde hace más de tres siglos persigue a los filósofos con su pregunta: ¿Por qué debo ser «justo» y cumplir los contratos, aun cuando ello sea

---

<sup>1</sup> Cualquier mérito que tenga el presente trabajo debe entenderse como una muestra de los beneficios que pueden resultar de empresas cooperativas. Las ideas aquí presentadas no hubieran sido concebidas sin las contribuciones de cada uno de los miembros del grupo de trabajo que en los años 1988/89 se reunió varias veces para discutir sobre el pensamiento de David Gauthier; mis «socios» en este grupo eran Manuel Atienza, Albert Calsamiglia, Ricardo Caracciolo, Ernesto Garzón Valdés y Jorge Malem. Sin embargo, el trabajo es también una prueba de solidaridad altruista: agradezco a Ernesto Garzón Valdés su heroico esfuerzo para eliminar una multitud de faltas lingüísticas y otras sustanciales.

irracional, ya que perjudica mis intereses? Evidentemente, la respuesta que dio Hobbes en su momento no era satisfactoria: el Tonto sigue preguntando. Hobbes sostenía que es racional cumplir con lo racionalmente acordado siempre que se tenga la garantía de que también las otras partes contratantes lo cumplirán. Según Hobbes, esta condición del cumplimiento universal podría satisfacerse, desde luego, sólo si existiera un poder dotado de medios coactivos suficientes. Hobbes se limitó, pues, a compensar la constatada falta del «fear of God» con el miedo frente a un poder político<sup>2</sup>. La amenaza de sanciones coactivas para el caso de la violación de un contrato cumple aquí dos funciones: por una parte, un individuo *A* puede estar seguro de que sus «socios» no tendrán ningún motivo para violar el contrato, ya que ello les resultaría costoso; queda así satisfecho el presupuesto hobbesiano para el cumplimiento del contrato por parte de *A*. Pero, por otra, cambia también -para decirlo en términos modernos- el resultado del cálculo de costos y beneficios que hace *A* de sus propias alternativas de comportamiento; en muchos casos, sólo así pierde atractivo la alternativa del comportamiento del polizón, porque *A* sabe que ya no podrá obtener beneficio alguno de su violación del contrato cuando los otros lo cumplen<sup>3</sup>.

El inconveniente de esta versión de una «solución política» a la inestabilidad de los contratos es que la racionalidad del cumplimiento de lo acordado -del comportamiento «justo»- resulta ser contingente. Depende de la existencia de circunstancias empíricas cuya creación y mantenimiento no son -justamente desde la perspectiva propuesta por Hobbes en su respuesta a la pregunta del Tonto- nada obvios<sup>4</sup>.

---

<sup>2</sup> Acerca de la concepción según la cual «la teoría política liberal de Hobbes se basa en la eliminación del miedo frente al “poder de los espíritus invisibles”, reforzando al mismo tiempo el miedo frente al “poder de los hombres”», cfr., entre otros, el trabajo reciente de *Blits*, 1989, esp. págs. 426 y sigs.

<sup>3</sup> Cfr. *Taylor*, 1987, pág. 136, para la distinción entre estas dos funciones: sostiene que, a pesar de que suele tomarse en cuenta sólo la segunda de ellas, «Hobbes obviamente no afirma simplemente que el Soberano logrará mantener la paz porque cada individuo le obedecerá sólo por miedo a las sanciones. Cree que ... cada uno le obedecerá porque el Soberano eliminará la única razón en contra del cumplimiento de los acuerdos, es decir, la “sospecha razonable” de que los demás no cumplirán los suyos». Taylor cita a H. L. A. Hart y B. Barry para interpretaciones similares. Desde luego, la primera función es suficiente -también según Taylor- sólo bajo determinadas circunstancias. Cfr. también las diferentes interpretaciones del dilema del polizón en *Hampton*, 1986.

<sup>4</sup> Cfr. con respecto al «problema de ordenamiento de Hobbes», entre muchos otros, *Kliemt*, 1987 y 1989.

Otra vía para llegar a una respuesta a la pregunta del Tonto ha sido la de admitir que, efectivamente, la justicia y la racionalidad, al menos a veces, dictan comportamientos distintos (por ejemplo, en lo que respecta al cumplimiento de los contratos), pero que ello no provoca conflicto alguno. Se puede pensar aquí básicamente en dos tipos de argumentos:

a) Sostener que no existe -y tampoco debería existir- *ninguna* relación entre racionalidad y moralidad, es decir, que la moralidad se basa totalmente en consideraciones extrarracionales (por ejemplo, metafísicas) y que tales criterios morales *sui generis* simplemente tienen prioridad frente a los criterios racionales.

Esta «solución teológica» difícilmente puede convencer a nuestro Tonto, ya que se trata, como es bien sabido, de un «Tonto racional» que no está dispuesto a creer en «dioses», de cualquier tipo que ellos sean.

b) Reconocer las discrepancias que existen entre lo racional y lo moral, pero sin inferir de ello una separación completa de los dos ámbitos. Por consiguiente, a diferencia de las soluciones propuestas por Hobbes y por los «teólogos», no habría aquí ninguna necesidad de decidirse por un principio de prioridad exclusiva en favor de la racionalidad o de la moralidad. Se insistiría, más bien, en la existencia de una relación entre los dos conceptos, asignándole a la racionalidad un papel importante en el ámbito de la moral; pero este papel se limita a la determinación del *contenido* de los comportamientos debidos -entre los que podría figurar, por ejemplo, el de «cumplir lo racionalmente acordado»-, y no se extiende a la constitución de los *motivos* para la realización del comportamiento debido. Por ello, podría ser debido, por ejemplo, el comportamiento racionalmente acordado, aun cuando no se pueda dar una «razón racional» para su cumplimiento en un caso particular. Según esta argumentación, si el Tonto carece de lo que podríamos llamar un «sentido de justicia», lo único que cabe decir es: peor para los demás; hay que cumplir lo racionalmente acordado justamente por haber sido racionalmente acordado, y por ninguna razón ulterior; dicho con otras palabras, hay que ser justo, no porque sea racional, sino simplemente porque es justo. Esta «solución moral» (en sentido estricto) parece ser la propuesta, entre otros, por Rawls, con su insistencia en el sentido de justicia.

Al «Tonto racional», por supuesto, no le satisfará ninguna de las soluciones propuestas: ni la «solución política» a la Hobbes (que, por cierto, da prioridad a la racionalidad, pero no logra establecer su coincidencia incondicionada con la moralidad) ni

la «solución moral» a la Rawls (que subordina la racionalidad a la moral) ni mucho menos la «solución teológica» (que no confiere a la racionalidad relevancia alguna en los asuntos de la moral).

La única solución que podría satisfacerlo plenamente sería aquella que contestara a su pregunta aduciendo que hay que ser «justo» -por ejemplo, hay que cumplir los contratos- porque ésta es la única actitud *racional*.

## II. LA RESPUESTA DE GAUTHIER

La fundamentación argumentativa de esta respuesta es justamente la tarea que se ha propuesto llevar a cabo David Gauthier con su libro *Morals by Agreement* (1986).

El argumento básico de Gauthier es que el Tonto de Hobbes no merece el nombre de «Tonto racional» -en el sentido de «hombre perfectamente racional, pero sin sentido de justicia»<sup>5</sup>. Intenta mostrar que, efectivamente, su racionalidad es deficiente -es decir, que es un Tonto sin más- y que no formularía su pregunta si realmente fuera racional.

Parte para ello, siguiendo a Hobbes y a Rawls, de una perspectiva contractualista. La elección de este punto de partida no es en modo alguno arbitraria; por el contrario, se infiere lógicamente del criterio de racionalidad: Sería ilusoria la esperanza de poder convencer a un presunto «Tonto racional» para que acepte como debidos comportamientos que no le benefician; pero, los comportamientos beneficiosos son justamente aquellos que una persona racional podría aceptar como contenido de un contrato. Por ello, la perspectiva contractualista abarca todo el campo de los comportamientos racionalmente aceptables, es decir, todo el campo relevante para la discusión con el Tonto.

Pero, mientras que Hobbes respondía al Tonto que racionalidad y moralidad coinciden *bajo ciertas condiciones externas*, y que no necesariamente coinciden cuando ellas no se dan, y Rawls sostenía que, en última instancia, no importa si el comportamiento racional y el moral coinciden o no, la respuesta de Gauthier implica que la moralidad coincide siempre con la racionalidad, es decir, que el comportamiento racional siempre es, al mismo tiempo, moral, sin que para ello se requiera la satisfacción

---

<sup>5</sup> Para una crítica a este tipo de «Tontos racionales», cfr. sobre todo el famoso artículo de Sen, 1982.

de ninguna condición adicional como no sea la racionalidad de todos los participantes en la interacción respectiva.

Más concretamente, con respecto al cumplimiento de los contratos, ello significa lo siguiente: En Hobbes, el cumplimiento de un contrato se vuelve racional a causa de los *costos externos* de la deserción; son ellos los que vuelven irracional la actitud del polizón, y sin ellos, el polizón sería -así se puede interpretar a Hobbes- perfectamente racional, aunque inmoral. En Rawls, donde no existe la amenaza de sanciones (al menos, no necesariamente), cumplir un contrato puede resultar menos beneficioso que la deserción «unilateral», pero ello es irrelevante, ya que sólo interesa el *status* moral de las acciones, y en este sentido el comportamiento del polizón es considerado siempre como inmoral. En cambio, a Gauthier, lo que le interesa es mostrar que el comportamiento del polizón, en un contexto de individuos racionales, *siempre* es irracional, y que, por el contrario, cumplir lo acordado *siempre* es racional; y como, también según Gauthier, cumplir los contratos es moral o «justo», racionalidad y moralidad coinciden.

Gauthier parte, efectivamente, de la concepción antropológica más pesimista imaginable: lo único que nos importa en la elección de nuestras acciones es -según él- el beneficio que podemos esperar de ellas en términos de la satisfacción de nuestras preferencias individuales. El hecho de que posiblemente esta premisa sea demasiado pesimista no afecta en nada la relevancia de la argumentación gauthieriana; por el contrario: si logra convencer a los empedernidos racionales y amorales, ¡cuánto más fácil será convencer a los que ya disponen de un sentido de justicia! Además, señala Gauthier, aun a estos últimos les conviene tener en cuenta, en sus interacciones con otras personas, sus propias preferencias e intereses y, si fuera necesario, revisar los términos de la interacción. «En efecto -dice-, suponemos que cierta medida de sociabilidad es característica de los seres humanos. Pero... [ésta] se convierte en fuente de explotación si lleva a las personas a tolerar instituciones y prácticas que, si no fuera por su sentido de compasión y afecto para los demás, les resultarían costosas. El pensamiento feminista nos ha aclarado definitivamente este mecanismo que es, quizás, la forma básica de toda explotación humana»<sup>6</sup>.

---

<sup>6</sup> Gauthier, 1986, pág. 11. Más claramente aún lo ha formulado Braybrooke 1987, pág. 753: «Hasta las personas vinculadas por estrechos lazos de afecto -amantes, matrimonios felices- probablemente harían bien en preguntarse de vez en cuando si la división de beneficios entre ellos toma adecuadamente en cuenta sus diferentes intereses.»

Vistas así las cosas, el programa de Gauthier consiste en mostrar:

1. Que la racionalidad a veces prescribe no sólo celebrar un contrato, sino también cumplirlo. Evidentemente, esto no vale para cualquier tipo de contratos, sino sólo para aquellos que satisfacen ciertas condiciones.

2. Que es racional celebrar un contrato si y sólo si también es racional su cumplimiento. Esto es equivalente a la afirmación de que para una persona racional siempre es racional cumplir los contratos<sup>7</sup>.

Con ello, Gauthier ya habría contestado de manera concluyente a la pregunta del Tonto. El Tonto, una vez convencido, deja de ser un «Tonto sin más» y, por fin, se convierte en un ser estrictamente racional. Pero Gauthier no se contenta con ello. Su objetivo último es convencer también a todos los demás de que este nuevo ser ya no merece la denominación maliciosa de «Tonto racional». Por ello, la última parte de su programa consiste en mostrar:

3. Que el comportamiento prescrito por la racionalidad satisface, al mismo tiempo, todos los requisitos necesarios y conjuntamente suficientes para poder hablar de un comportamiento moral.

#### a) *El concepto de racionalidad*

Para fundamentar su afirmación en el sentido de que la racionalidad del Tonto es deficiente, Gauthier tiene básicamente dos posibilidades. Puede rechazar por completo el concepto de racionalidad del Tonto y aducir que los criterios de racionalidad deberían ser otros; o puede vencer al Tonto «con sus propias armas», haciendo suyos precisamente los mismos criterios del Tonto y mostrándole que no los aplica coherentemente. Desde luego, la primera alternativa sería más fácil: no es precisamente por escasez de conceptos que tenemos tantos problemas con la racionalidad<sup>8</sup>; pero esta alternativa conduciría a una polémica con alto riesgo de esterilidad, ya que en el ámbito de las definiciones

---

<sup>7</sup> Obviamente, ello no vale para una persona que no es (perfectamente) racional, ya que en este caso puede haber contratos cuyo cumplimiento sería todo menos racional.

<sup>8</sup> Cfr., por ejemplo, la lista con 21 significados diferentes del término «racionalidad» que presenta *Hans Lenk*, 1986, págs. 106 y sigs.; cfr. también *Harsanyi*, 1985, y *Mendola*, 1987.

no es posible aducir criterios concluyentes, como los de verdad o falsedad.

Felizmente, Gauthier opta por la vía más difícil y más interesante, al utilizar un concepto básico de racionalidad similar al del Tonto. Se trata de aquella racionalidad que «dicta a cada uno [procurar] su propio bien»<sup>9</sup>: Una persona es racional en la medida en que elige aquellas acciones que maximizan su utilidad esperada, donde esta utilidad depende exclusivamente de sus preferencias ponderadas (*considered*) que, a su vez, derivan del valor -subjetivo y relativo- que el individuo asigna a cada uno de los posibles estados de cosas alternativos<sup>10</sup>. Por consiguiente, la elección de una acción no maximizadora de utilidad debe ser considerada como irracional. Se definen así límites muy estrechos para las restricciones racionalmente aceptables a la maximización de utilidad. A primera vista, hasta parece que la racionalidad y la restricción a la maximización de utilidad se vuelven *per definitionem* recíprocamente excluyentes. Sin embargo, no es difícil percibir que esta impresión es falsa: Si la utilidad máxima que puedo sacar de mis ahorros actuales, dadas mis preferencias ponderadas, consiste en una semana de vacaciones en la playa, y si decido no realizar esta utilidad, ello restringe, sin duda alguna, la maximización de mi utilidad; si, además, gasto mis ahorros en alguna otra cosa y así realizo una utilidad menor que la que hubiera obtenido de la semana en la playa, la restricción a la maximización es claramente irracional. Pero, si renuncio a la realización de la utilidad máxima y de cualquier utilidad menor -porque sólo así puedo tener la esperanza de disponer algún día de los recursos necesarios para comprarme aquella casa en la playa con la que estoy soñando desde siempre y cuya utilidad (esperada) para mí será mucho mayor que la de una sola y corta visita a la misma playa- entonces la restricción, lejos de ser irracional, no es otra cosa que una estrategia totalmente adecuada para una maximización de utilidad a largo plazo. En este sentido, mi disposición a cumplir disciplinadamente con las restricciones a la maximización de utilidad que me impongo a mí misma es racional porque, en última instancia, sirve para elevar el máximo de utilidad «esperable», alcanzable -con lo cual, desde luego, cambia también el cálculo de utilidad esperada.

Lo mismo vale en el contexto de la interacción con los demás: «En tanto maximizador racional de utilidad, debo cumplir

---

<sup>9</sup> Cfr. *Hobbes, loc. cit.*

<sup>10</sup> Cfr., sobre todo, *Gauthier, 1986, capítulo II.*

con una restricción si y sólo si puedo dar para ello una razón maximizadora de utilidad. Ahora bien, una razón directa obviamente no es posible... Por ello, es racional adoptar aquellas disposiciones, con respecto al cumplimiento de principios restrictivos, que los demás consideran ventajoso recompensar»<sup>11</sup>.

b) *La superación del problema del polizón*

La posibilidad de que una restricción a la maximización directa, a través de la obtención de beneficios que surgen de la restricción misma, sirva para la maximización indirecta de utilidad, es decir, la posibilidad de una restricción *racional* a la maximización de utilidad, es la idea clave para la argumentación de Gauthier en contra del Tonto.

Gauthier señala que hay «toda una gama de situaciones -entre ellas, en primer lugar, las del “dilema de los prisioneros” en forma general -en las que estrategias que exigen un comportamiento no maximizador realmente maximizan la utilidad y son, por ello, racionales»; ello es así porque tales estrategias introducen un cambio en «las probabilidades de que el agente se encuentre en ciertas situaciones, frente a ciertas alternativas. Sólo si entendemos esto, podemos comprender las verdaderas características y la complejidad de la racionalidad maximizadora de utilidad»<sup>12</sup>. La probabilidad, por ejemplo, de que algún día me encuentre en una situación que ofrezca la alternativa directa de comprarme aquella casa en la playa aumenta en la medida en que persigo, durante algún tiempo, una estrategia indirecta, acumulando ahorros, en lugar de satisfacer mis preferencias de cada momento. Este mecanismo de las estrategias indirectas de maximización adquiere más importancia aun cuando se trata de beneficios realizables sólo mediante la colaboración de otras personas, es decir, cuando las situaciones en las que habré de encontrarme, y las alternativas que habrán de presentármese, ya no dependen sólo de mi propia disciplina en el cumplimiento de las restricciones que me impongo por mi propio bien, sino también

---

<sup>11</sup> Gauthier, 1988b, pág. 183; cfr. también Gauthier, 1988, pág. 13, donde repite: «Es racional disponerse a aceptar ciertas restricciones a la maximización directa, en la elección y el comportamiento, si y sólo si una disposición tal maximiza la utilidad esperada.»

<sup>12</sup> Gauthier, 1984, pág. 489.



de la disciplina de otras personas en el cumplimiento de las restricciones que ellas mismas se han impuesto.

Ciertamente, entre personas racionales no se puede contar con el cumplimiento voluntario, espontáneo, de restricciones que no aumentarán la utilidad máxima esperable para quien las cumpla. Por ello, en el caso de personas racionales, no vinculadas por lazos de afecto con las demás, el cumplimiento de las restricciones a la maximización directa de utilidad en aras de alguna empresa cooperativa es esperable, cuando más, si se trata de empresas mutuamente beneficiosas.

Pero aun en este caso, queda por resolver el viejo dilema de los prisioneros -víctimas de su racionalidad maximizadora: ¿Cómo evitar que los beneficios potenciales de la cooperación queden sin realizarse porque cada uno, sospechando que está frente a un polizón, ante la alternativa del riesgo de ser una víctima «ingenua» de la explotación, por un lado, y de los costos (de oportunidad) seguros, pero claramente menores, de la irrestricta persecución de la maximización directa de utilidad, por otro, elige esta última vía? ¿Cómo demostrarle plausiblemente al Tonto que justamente esta elección sería irracional?

Al reflexionar acerca de la *simetría* de las situaciones de los cooperadores potenciales, Gauthier encuentra la solución del problema. Resumida en muy pocas palabras, es básicamente la siguiente: En una empresa cooperativa, un maximizador directo de utilidad siempre *debe* (racionalmente) convertirse en polizón, es decir, aprovecharse de la cooperación de los demás y privarlos de los beneficios que habían esperado de sus respectivas restricciones a la maximización directa de sus utilidades. Justamente por ello, una persona racional jamás cooperará con alguien de quien se sabe o se sospecha que es un polizón; por consiguiente, el maximizador directo de utilidad quedará siempre al margen de todo beneficio realizable sólo mediante un esfuerzo común. Así pues, la única posibilidad que tiene una persona de convertirse en un socio aceptable para los demás -y asegurarse así alguna participación en los posibles beneficios cooperativos- consiste en convencerlos de que se encuentran frente a una persona que *no* se convertirá en polizón, es decir, que no persigue la maximización directa de su utilidad. Desde luego, sería irracional renunciar incondicionadamente a la maximización directa, porque ello equivaldría a una invitación a la explotación; la renuncia debe hacerse con la reserva de que ella vale exclusivamente para la interacción con personas igualmente dispuestas. En palabras de Gauthier: «La disposición de cumplir los contratos, *en condiciones de seguridad suficiente*, sin recurrir a consideraciones

acerca de la maximización directa de utilidad, lo convierte a uno en socio elegible para la cooperación beneficiosa y, por ello, es a su vez beneficiosa»<sup>13</sup>.

Pero esto, afirma Gauthier, no basta todavía para asegurar que efectivamente se realice una cooperación mutuamente beneficiosa. Una condición necesaria para que un individuo sea elegido como socio en una empresa cooperativa concreta es, por cierto, que renuncie a la maximización directa de su utilidad, pero ello no es una condición suficiente. Pues, una vez realizada la cooperación, siempre surge la cuestión de la distribución del beneficio resultante entre los participantes. Para cada uno de ellos, es racional aportar su parte del esfuerzo común sólo si puede estar seguro de que, después, recibirá una parte «adecuada» del beneficio comúnmente obtenido. Por lo tanto, para que se obtenga algún beneficio, es necesario que los socios potenciales no sólo se reconozcan recíprocamente como personas que, bajo ciertas condiciones, están dispuestas a renunciar a la maximización directa de su utilidad, sino que, además, se pongan de acuerdo acerca de los términos concretos de la distribución del beneficio de su eventual cooperación. A este acuerdo se llega a través de una negociación en la que cada uno de los participantes tiene que tomar en cuenta dos aspectos fundamentales: 1) que su propia racionalidad (indirectamente) maximizadora le exige, por supuesto, aceptar sólo aquellos acuerdos que le aseguran un beneficio «neto», es decir, una parte del beneficio global de la cooperación cuyo valor supera el de los costos que el mismo individuo asume en la cooperación; más aún, le exige también que no se «explote» a sí mismo -aceptando una parte que, si bien significa un beneficio neto, es menor que la que podría obtener de sus socios racionales- sino que reclame para sí ni más ni menos que la máxima parte del beneficio global que ellos deberían estar racionalmente dispuestos a concederle; y 2) que todos los demás tienen exactamente las mismas preocupaciones. Por ello, si se está realmente interesado en la obtención de algún beneficio cooperativo, no puede ser racional pedir de los socios potenciales concesiones «exageradas».

La cuestión ahora es saber cómo reconocer el punto en que una exigencia deja de ser racional y se convierte en exagerada. En el capítulo V de su libro, Gauthier desarrolla al respecto una

---

<sup>13</sup> *Gauthier*, 1986, pág. 162 (subrayado R. Z.); con ello, Gauthier sostiene que, en todo caso, para una persona racional sería irrelevante la segunda de las más arriba mencionadas funciones de la amenaza de sanciones en la propuesta hobbesiana.

argumentación detallada y sofisticada, para mostrar que el resultado de una negociación entre personas racionales, con el fin de llegar a un acuerdo sobre la distribución de los beneficios esperados de una empresa cooperativa, consiste en una distribución tal que no exista ninguna otra distribución también racionalmente aceptable por todos y en la que la concesión relativa máxima sea menor<sup>14</sup>. Esto significa, en la terminología de Gauthier, que la negociación racional de un acuerdo acerca de los términos de una empresa cooperativa está guiada por el «Principio de la concesión relativa minimax». La persona racional, según Gauthier, no entra en acuerdos que no satisfacen este principio; pero, si ha celebrado un acuerdo que lo satisface, es también racional cumplirlo: «El mismo principio, el de la concesión relativa minimax, sirve a las personas racionales, tanto para llegar a un acuerdo como para cumplirlo, es decir, sirve de principio de elección tanto en la negociación como en la cooperación... La negociación es una actividad directamente maximizadora, que conduce a un acuerdo sobre una estrategia común. La restricción aparece tan sólo en la interacción cooperativa misma que exige la adhesión a esta estrategia...»<sup>15</sup>

Gauthier ha completado así su respuesta a la pregunta del Tonto. Parece haber mostrado que una persona racional, si tiene la oportunidad de negociar un acuerdo acerca de los términos de una empresa cooperativa mutuamente beneficiosa, debe aplicar el Principio de la concesión relativa minimax porque sabe que ello maximiza su utilidad: si insiste en pedir más, la cooperación no se llevará a cabo, ya que sería irracional aceptarla por parte de los otros, y no habrá beneficio alguno; y si pide menos, renuncia innecesaria y, por ello, irracionalmente a una parte de utilidad que le sería posible obtener. Parece haber mostrado también que una persona racional debe disponerse a cumplir los acuerdos, porque de otro modo sus congéneres (igualmente racionales)

---

<sup>14</sup> Por ejemplo, en una negociación entre tres personas A, B y C: si hay una distribución posible (es decir, racionalmente aceptable para cada una de las tres y que, por ello, permite la cooperación) en la que A debería hacer la concesión más grande y conceder el 75 por 100 a las otras, contentándose con un 25 por 100 para sí misma, y si hay otra distribución posible en la que nadie debería hacer una concesión mayor del 66 por 100 del beneficio total (cada una contentándose con el 33 por 100 para sí misma), entonces es claramente irracional para A aceptar la primera alternativa. Desde luego, para evitar los problemas conocidos de la comparación intersubjetiva de utilidades, Gauthier habla siempre de concesiones relativas (relativas a la utilidad -subjetiva- que el total del beneficio cooperativo significaría para cada uno respectivamente).

<sup>15</sup> *Gauthier*, 1986, págs. 146 y 151.

le negarán la participación en cualquier acuerdo, con lo que pierde toda posibilidad de beneficiarse de empresas cooperativas -sea como polizón, sea como un socio entre otros.

c) *La moralidad de la racionalidad*

Finalmente, Gauthier pretende mostrar que el Tonto, si aplica coherentemente su propio concepto de racionalidad y se dispone a cumplir los acuerdos racionales, no sólo es racional sino también moral y, por lo tanto, ya no merece el nombre de «Tonto» ni en uno ni en otro sentido.

Tiene que mostrar, para ello, que el cumplimiento racional de acuerdos racionales *coincide* con la moralidad. Aquí cabe señalar, desde luego, que Gauthier no pretende mostrar que la moralidad de que está hablando coincide, a su vez, con «la moralidad que aprendimos de nuestros padres y compañeros, nuestros curas y profesores»<sup>16</sup>. Sobre todo, la moralidad gauthieriana -dado que la racionalidad con la que debe coincidir está, como vimos más arriba, estrechamente vinculada con una evaluación subjetiva y relativa de las cosas- difícilmente podría adecuarse a una concepción de la moralidad «como restricción con respecto a los objetos deseados, impuesta por nuestro conocimiento de una bondad objetiva en el mundo. Según este punto de vista ... [una] persona racional y moral intentaría maximizar la satisfacción de sus deseos moralmente restringidos»; en cambio, según Gauthier, las únicas restricciones justificables son las restricciones racionales, y la fuerza de las prescripciones morales «deriva totalmente de su papel en la superación del problema estructural de la interacción presentada en el dilema de los prisioneros ... Todo lo demás sería sólo un *addendum* mítico que, una vez comprendido, volvería a la moralidad tan impotente como ya lo es la religión»<sup>17</sup> o, para utilizar la expresión más lírica de González Ballesteros: «Lo demás es música de fandangos celestiales apócrifos»<sup>18</sup>.

Según este punto de vista, pues, «la base de la moralidad» no puede ser otra que un «acuerdo unánime entre personas racionales» acerca de «restricciones mutuas ... mutuamente ventajosas»<sup>19</sup>. Sin embargo, como afirma el propio Gauthier, la moralidad

---

<sup>16</sup> *Ibid.*, pág. 6.

<sup>17</sup> Gauthier, 1988a, págs. 175 y sig.

<sup>18</sup> Teodoro González Ballesteros, «El cumplimiento del pacto», en *El País* del 6 de nov. 1989.

<sup>19</sup> Gauthier, 1988b, pág. 10.

«no debe ser entendida como una restricción a la satisfacción de preferencias no racionales, que surge sólo de la racionalidad. Más bien, un agente racional es aquel que actúa para lograr la satisfacción máxima de sus preferencias, y la moralidad es una restricción a su manera de actuar que surge de los efectos de su interacción con otros agentes»<sup>20</sup>.

El significado fundamental de «moralidad», según Gauthier, es la prescripción de restricciones a la libre persecución del autointerés. La racionalidad -como acabamos de ver- también exige restricciones de este tipo; por ello, es básicamente posible una coincidencia de racionalidad y moralidad. Pero, mientras que las restricciones racionales a la maximización directa se producen *por razones maximizadoras*, la moralidad, *per definitionem*, consiste en restricciones. Ahora bien, para poder contestar a la pregunta de si las restricciones dictadas por la racionalidad coinciden con las prescritas por la moralidad, hay que saber cuáles son los criterios que hacen de una restricción una restricción moral. Según Gauthier, los únicos criterios necesarios para que una restricción sea moral son los de la *autoimposición* voluntaria y la *imparcialidad*<sup>21</sup>. En el contexto de personas racionales, el primero de los dos criterios está satisfecho si y sólo si se trata de una restricción racional. Por ello, las restricciones *morales* son justamente aquellas restricciones *racionales* que son también *imparciales*; el ámbito de la moralidad es idéntico al ámbito de las restricciones racionales, imparciales<sup>22</sup>.

Con respecto a la imparcialidad, el argumento de Gauthier sostiene que las restricciones racionales *siempre* son imparciales, es decir, el resultado de un acuerdo racional es siempre *equitativo*. Por lo tanto, la exigencia de imparcialidad no añade nada a la exigencia de racionalidad, es decir, la racionalidad y la moralidad coinciden plenamente.

---

<sup>20</sup> *Ibid.*, pág. 11.

<sup>21</sup> Gauthier, 1988a, pág. 177: «[U]na restricción moral se caracteriza por dos propiedades: es una restricción interna que opera a través de la voluntad o toma de decisiones del agente, y opera de una manera que satisface algún estándar de imparcialidad entre las personas»; cfr. también Gauthier, 1988b, págs. 17 y sig., donde afirma que «sería erróneo considerar las restricciones racionalmente acordadas a la maximización directa de utilidad como una moralidad en sí, en lugar de como un sustituto a la moralidad, si no fuera por la coincidencia de su contenido con nuestras opiniones morales preteóricas. Y la coincidencia se debe -según creo- al hecho de que el consenso hipotético exige la imparcialidad».

<sup>22</sup> Cfr. Gauthier, 1986, pág. 6: «Pretendemos mostrar que existen restricciones racionales, y que estas restricciones son imparciales. Identificamos, pues, la moralidad con ellas.»

La fundamentación de este argumento presenta dos aspectos diferentes. Primero, Gauthier sostiene que la negociación racional, guiada por el Principio de la concesión relativa minimax, siempre es imparcial y equitativa. Ello parece plausible ya que el Principio mismo se infiere de la reflexión acerca de las preocupaciones simétricas de todas las partes contratantes: ninguna de ellas aceptaría ni desventajas particulares para sí misma ni ventajas particulares para alguna de las otras. Pero si no se admiten «privilegios» o «excepciones», entonces la negociación es imparcial, es decir, equitativa. Pero, segundo, surge ahora un problema que todavía no ha sido mencionado. En la negociación, el Principio de la concesión relativa minimax implica exclusivamente la imparcialidad de la distribución del beneficio de la cooperación. Se refiere, por así decirlo, sólo al futuro. Por lo tanto, el resultado de una negociación imparcial conduce a un estado de cosas equitativo sólo si parte ya de un estado de cosas equitativo. La negociación imparcial no puede corregir «injusticias»<sup>23</sup> del pasado; especialmente, no da lugar a ninguna redistribución.

La imparcialidad como requisito de la moralidad exige, obviamente, que en ningún momento -tampoco antes de la celebración del primer acuerdo- exista una distribución injusta de bienes, es decir, una distribución en la que alguna persona se haya apropiado de lo que le corresponde a otra. Ahora bien, lo que le corresponde a una persona -prescindiendo de lo que ha obtenido a través de negociaciones imparciales- es todo aquello que, con sus propios esfuerzos, hubiera obtenido sin ninguna interacción con otras personas. Ello significa especialmente que sólo le corresponde aquello que no ha obtenido a través de una interacción perjudicial para los demás y que hubiera podido evitar sin perjudicarse a sí misma. Por lo tanto, según Gauthier, la imparcialidad implica la observancia de una variante de la «cláusula cautelar de Locke» que «le prohíbe a uno mejorar su propia situación mediante una interacción que empeora la situación de otra persona. Ello expresa ... la idea subyacente de no aprovecharse de los demás»<sup>24</sup>. Dicho de otra manera, la moralidad exige que se reconozcan los *derechos* de cada persona a ciertos bienes

---

<sup>23</sup> Cfr. *ibid.*, pág. 150: «La justicia es la disposición a cumplir ... una restricción imparcial.»

<sup>24</sup> *Ibid.*, pág. 205; la derivación de esta «cláusula cautelar de Gauthier» y la demostración de su imparcialidad y racionalidad son el tema del cap. VII del libro. Para la versión «original» de la cláusula, cfr. *Locke*, 1689, cap. V (Of property), § 27.

y que la persecución de la maximización de utilidad se restrinja a aquello que puede obtenerse sin aprovecharse de los demás y violar así sus derechos.

Como ya se ha visto, por lo que respecta a la racionalidad, Gauthier sostiene lo siguiente: La celebración y la disposición al cumplimiento de un acuerdo son racionales si y sólo si se tiene la garantía de un resultado imparcial. Pero la imparcialidad depende, tanto del Principio de la concesión relativa minimax como de la cláusula cautelar. Por lo tanto, «la disposición al cumplimiento [de los acuerdos imparciales] incluye la disposición a la aceptación de esta cláusula como restricción a la interacción natural, en la medida en que se tiene la esperanza de ingresar en la sociedad, ... en las prácticas cooperativas»<sup>25</sup>.

Con ello, Gauthier ha llegado al final de su argumentación que concluye con la afirmación de que «la justicia, es decir, la disposición a no aprovecharse de sus congéneres, es la virtud apropiada para la cooperación y es voluntariamente aceptada por personas igualmente racionales»<sup>26</sup>.

### III. OTRA VEZ, PREGUNTAS

La extensa respuesta elaborada por David Gauthier a la pregunta del Tonto, en el sentido de que para ser «justo» sólo sería necesario ser racional, ha sido objeto ya de un gran número de comentarios críticos. En su mayoría, estas críticas se refieren a los detalles de la argumentación «técnica» de Gauthier e intentan mostrar que sus inferencias no siempre son válidas<sup>27</sup>. Este no es el lugar para entrar en esos detalles y repetir lo que ya se ha dicho al respecto. Más interesante parece analizar si la argumentación de Gauthier -prescindiendo de la necesidad de corregirla en algunos puntos- básicamente tiene alguna oportunidad de

---

<sup>25</sup> Gauthier, 1986, pág. 226.

<sup>26</sup> *Ibid.*, pág. 232.

<sup>27</sup> Cfr., aparte de los ensayos críticos en Frankel Paul et al, 1988, sobre todo la discusión muy detallada de Kraus/Coleman, 1987, que sostienen la implausibilidad de algunas de las premisas y la incoherencia lógica de algunas de las conclusiones básicas de Gauthier; además, Hegselmann, 1988, quien señala algunos problemas en los cálculos que presenta Gauthier para mostrar que bajo ciertas condiciones es más beneficioso ser un maximizador restringido que un maximizador directo; como así también la reseña de Nelson, 1988, con objeciones empíricas y conceptuales en contra de la idea de que sea posible adoptar voluntariamente una disposición a ser un maximizador restringido, parcialmente reconocible (*traslúcido*).

éxito, es decir, de poner fin de una vez para siempre a la preguntona actitud del Tonto. Para ello, me limitaré en lo que sigue a formular tres nuevas preguntas que, desde el punto de vista de un presunto Tonto, se le pueden hacer a Gauthier.

1. *¿Es realmente racional aceptar la cláusula cautelar?*

La argumentación de Gauthier concluye, como acabamos de ver, con la afirmación de que una persona racional debe aceptar la versión gauthieriana de la Cláusula cautelar de Locke que prohíbe aprovecharse de los demás. Gauthier fundamenta esta tesis de la siguiente manera: En el primer paso (en los caps. V y VI), muestra que una persona racional celebra y cumple sólo aquellos contratos que prometen un resultado imparcial y que son, por ello, equitativos. En el segundo paso (en el cap. VII), aduce que la imparcialidad implica dos cosas: el Principio de la concesión relativa minimax (PCRM), con respecto a la negociación, y la cláusula cautelar (CC), con respecto a la situación inicial de la negociación. De ello infiere, tercero, que una persona racional debe aceptar la CC. El problema con esta conclusión es que la imparcialidad cuya racionalidad Gauthier había mostrado en el primer paso, se refería exclusivamente a la imparcialidad de la negociación, es decir, sólo a aquel aspecto de la imparcialidad «total» captado por el PCRM. Por lo tanto, el primer paso no contribuye nada a la conclusión de que es racional aceptar la CC<sup>28</sup>.

---

<sup>28</sup> Gauthier, 1986. En el cap. VI (pág. 178), define el «cumplimiento estrecho» (*narrow compliance*) como la disposición a «cooperar de una manera tal que, si todos la adoptan, proporciona resultados aproximadamente óptimos y equitativos». Luego, explica (pág. 179): «En el capítulo siguiente, extenderemos la concepción del cumplimiento estrecho para que incluya no sólo la satisfacción de la concesión relativa minimax, sino también la de algún *standard* de equidad para la posición inicial de la negociación.» Finalmente, en el cap. VII, el argumento reza (pág. 226): «Una persona dispuesta al cumplimiento estrecho compara el beneficio que se puede esperar de la cooperación con el que se podría esperar de un resultado óptimo y equitativo, y sólo cumple con una estrategia común si aquél se aproxima a éste. Un resultado es equitativo si satisface los *standards* de la imparcialidad que consisten, como hemos mostrado, en la cláusula cautelar y en el Principio de la concesión relativa minimax. *Por lo tanto*, una persona dispuesta al cumplimiento estrecho supone que los demás adoptan -y consideran racional adoptar- la cláusula como una condición para la cooperación. Pero entonces, dada la igualdad de racionalidad, ella misma también debe considerar racional adoptar la cláusula como condición para la cooperación. *Por lo tanto*, la disposición al cumplimiento



La argumentación de Gauthier tiene, pues, una «laguna» justamente en uno de los puntos decisivos de su programa; hasta que se logre llenarla con argumentos que no den lugar a dudas con respecto a la racionalidad de aceptar la CC -es decir, de aceptar ciertos derechos y con ello, por lo menos, ciertos deberes negativos, sin recurrir a argumentos morales- no se habrá mostrado la coincidencia de la racionalidad con la moralidad. El hecho de que justamente en este punto central, falle la «perfección técnica» de Gauthier, generalmente reconocida hasta por sus críticos<sup>29</sup>, no permite alentar una actitud optimista con respecto a una posible superación del problema.

## 2. ¿Es realmente racional la maximización restringida?

La segunda pregunta plantea un problema más fundamental. Si Gauthier, con su advertencia de que no es racional la maximización directa sino la indirecta, no quiere decir nada más que, en general, «no es rentable» tomar en cuenta sólo las perspectivas inmediatas, sin pensar en las consecuencias más remotas de las acciones, obviamente no nos dice nada nuevo. El problema de la miope persecución de la satisfacción inmediata de nuestras preferencias, en detrimento de la consideración ponderada de las consecuencias a medio y largo plazo, fue tratado ya con toda claridad, por ejemplo, por David Hume<sup>30</sup>. Más recientemente, en 1979, Jon Elster escribió que «esperar y servirse de estrategias indirectas son características centrales de la toma de decisión humana»; propone, por ello, «considerar al hombre como una máquina globalmente maximizadora -una caracterización que se

---

estrecho incluye la disposición a aceptar la cláusula...» (subrayado R. Z.). Aquí, casi imperceptiblemente, en el curso de la argumentación, se cambia la noción de imparcialidad con el cambio del significado del término «resultado»: primero, se habla del resultado de una negociación; luego, del resultado del conjunto de interacciones que hasta un momento determinado se han llevado a cabo.

<sup>29</sup> Cfr. Braybrooke, 1987, pág. 751.

<sup>30</sup> Cfr. Hume, 1739/40, Libro III, Parte II, Sección VII (Sobre el origen del gobierno), donde se refiere a la «debilidad de la naturaleza humana» que lleva a la «propensión violenta de preferir lo contiguo a lo remoto» y dice: «Cuando consideramos los objetos desde la distancia, ... siempre preferimos aquello que sea en sí mismo preferible, sin considerar su situación y circunstancias. De ello, surge lo que, en un sentido impropio (*improper*), llamamos la *razón*, que es un principio frecuentemente contradictorio a aquellas propensiones que se manifiestan al acercarse el objeto.»

remonta a Leibniz; según él, el hombre es el único ser capaz de “reculer pour mieux sauter”<sup>31</sup>. Parece que el pensamiento de Gauthier, efectivamente, va en la misma dirección cuando, por ejemplo, observa -en tono de objeción- que «Hobbes nunca reconoce explícitamente que es motivacionalmente posible adoptar, por razones egoístas, una regla que prescribe acciones no egoístas»; su objetivo es justamente mostrar los motivos que «le facilitan a una persona adoptar una regla ventajosa, al precio de cometer actos particulares perjudiciales»<sup>32</sup>.

Sin embargo, una cosa es sostener que *muchas veces* la maximización indirecta es más beneficiosa que la directa, y otra que ello *siempre* sea así. Sólo si se afirma lo segundo, se puede también sostener que es racional adoptar la disposición a la maximización restringida, sin que en cada caso concreto se calculen los costos y beneficios. Si se afirma lo primero, por el contrario, no sería racional esta disposición: En cada caso concreto, una persona racional habrá de hacer un cálculo para saber si los posibles beneficios «netos» de la deserción (tomando en cuenta los costos de la consecuencia de que así se pierde toda oportunidad futura de obtener beneficios cooperativos) no superarían quizás los beneficios esperables, aun en el mejor de los casos, de las posibles empresas cooperativas futuras. En este sentido, la estrategia más beneficiosa -y, por ello, racional- sería posiblemente la simulación de la disposición a la maximización restringida,

---

<sup>31</sup> Elster, 1987 (publicación del original 1979), cap. I (Vollständige Rationalität: Jenseits gradualistischer Entwicklung), pág. 41, donde define: «Una máquina posee la capacidad de esperar si puede decir No a una mutación ventajosa para más tarde poder decir Sí a otra más ventajosa aún»; y «[u]na máquina posee la capacidad de adoptar estrategias indirectas si puede decir Sí a una mutación perjudicial para más tarde poder decir Sí a otra más ventajosa». Más adelante, Elster recuerda explícitamente la distinción fundamental entre la elección de una acción y la de una estrategia («un procedimiento para elegir»; pág. 43), para luego -cap. II (Unvollständige Rationalität: Odysseus und die Sirenen), pág. 96- pronunciarse partidario de la opinión de *Pascal* de que la racionalidad debería guiar solamente esta última, es decir, que una vez hecha la elección racional de una estrategia, las acciones deben seguir «automáticamente», sin que se considere a su vez, en cada caso, su racionalidad. Finalmente, sostiene que la idea central de *Descartes* era «que la modificación del comportamiento puede lograrse a través de una modificación autocontrolada (estratégica) del carácter» (pág. 86). Con todo, se muestra más bien escéptico acerca del alcance de la racionalidad estratégica en la realidad, debido a «la exigencia muy rigurosa de información» que ella supone (pág. 86), y concluye que «no tanto la falta de voluntad cuanto la falta de información impide la racionalidad perfecta» (pág. 93).

<sup>32</sup> Gauthier, 1987, págs. 287 y sig.

porque ella permite obtener todos los beneficios cooperativos posibles hasta el momento en el que se presenta la ocasión para la «deserción maximizadora»<sup>33</sup>. El propio Gauthier admite que «es una cuestión empírica saber si la simulación es lo mejor»<sup>34</sup>. Pero si es así, entonces es también una cuestión empírica -y no una mera cuestión de inferencia lógica desde premisas sobre la racionalidad y la estructura general de las interacciones humanas saber si es racional renunciar a la simulación y adoptar sinceramente una disposición de maximización restringida. Por lo tanto, la relación entre la equidad de un contrato y la racionalidad de su cumplimiento, sostenida por Gauthier, es más que dudosa: «...el hecho de que una regla sea considerada como equitativa ... no asegura, *per se*, la voluntad de cumplir esta regla. El problema del cumplimiento resulta del hecho de que hay beneficios potenciales en la deserción. La existencia o inexistencia de tales beneficios no depende *per se* de las propiedades de la equidad»<sup>35</sup>.

Sería extraño que el pensamiento de Gauthier realmente se limitase a una concepción de la racionalidad maximizadora tan fácilmente refutable. Y, en efecto, hay indicios de que sus reflexiones acerca de la racionalidad estratégica implican algo más. Gauthier sostiene que la noción «tradicional» de la racionalidad -utilizada por la teoría económica, la teoría de la decisión o la de los juegos- es falsa. Según él, la propia maximización de utilidad no debe referirse a rendimientos particulares (*performances*), sino a estrategias (*policies*)<sup>36</sup>. La diferencia le parece decisiva: «La maximización es la meta del maximizador de estrategias, y el destino del maximizador de rendimientos»<sup>37</sup>. Por ello, en el último párrafo de su libro, rechaza explícitamente la concepción del hombre de Elster en el sentido de la «máquina globalmente

---

<sup>33</sup> Cfr., por ejemplo, *Parfit*, 1987, pág. 18.

<sup>34</sup> *Gauthier*, 1987, pág. 297.

<sup>35</sup> *Vanberg/Buchanan*, 1989, pág. 55; en este contexto, los autores se refieren (*ibid.*, nota 11) explícitamente a Gauthier, aduciendo que «[l]a teoría de David Gauthier en *Morals by Agreement* parece basarse en parte en la suposición de que existe algún vínculo directo entre la equidad y el cumplimiento». Cfr. también *McClennen*, 1989, pág. 12 (n. 21), quien, a su vez, sostiene que «un sentido de la inequidad de un esquema por parte de algunos de los participantes no es una condición suficiente para su inestabilidad. Y sin duda, tenemos que admitir que un sentido de la inequidad de un esquema tampoco es una condición necesaria para la deserción unilateral...»

<sup>36</sup> Cfr., por ejemplo, *Gauthier*, 1986, pág. 182: «Identificamos la racionalidad con la maximización de utilidad en el nivel de las disposiciones para elegir.»

<sup>37</sup> *Gauthier*, 1984, pág. 488.

maximizadora», ya que aquí la maximización se refiere a los rendimientos (globales)<sup>38</sup>.

La concepción de la racionalidad como maximización de estrategias es justamente lo que le permite a Gauthier rechazar los argumentos, más arriba presentados, en contra de la racionalidad general de una disposición a la maximización restringida. Así, puede afirmar que «la maximización restringida no es la maximización directa en su disfraz más eficiente ... El maximizador restringido no razona más eficientemente acerca de cómo maximizar su utilidad, sino que razona de otra manera»: según Gauthier, el maximizador directo cumple un contrato sólo cuando ello le proporciona más beneficios que *cualquier* otro comportamiento (lo que explica la frecuencia de la deserción), mientras que el maximizador restringido cumple sus contratos siempre que el resultado (global) de la cooperación mutua de todos los socios sea óptimo y equitativo y si, al mismo tiempo, el resultado que el individuo puede esperar para sí mismo es mejor que el resultado esperado sin cooperación alguna<sup>39</sup>.

Aquí cabe señalar, desde luego, que el maximizador restringido así entendido parece padecer cierta presbicia en el sentido de que ya no logra enfocar aquellas alternativas que le proporcionarían un beneficio directo aún mayor al de la cooperación -deficiencia en la racionalidad justamente contraria a la más arriba mencionada miopía de los que no logran tomar en cuenta las consecuencias remotas de sus acciones.

Pero, parece haber un argumento aún más fuerte en contra de la concepción de la racionalidad como maximización de utilidad referida sólo a estrategias: El propio Gauthier advierte que no tendría sentido hablar de una «restricción» al comportamiento si la deserción no fuera beneficiosa (porque entonces no sería necesaria ninguna «restricción»)<sup>40</sup>. Ahora bien, si se entiende la racionalidad como maximización de la utilidad con respecto a resultados (rendimientos), entonces es fácil comprender que la maximización indirecta -la renuncia a beneficios inmediatos en aras de un mayor beneficio futuro- implica una restricción. Pero si los resultados (rendimientos) particulares ya no tienen

---

<sup>38</sup> Gauthier, 1986, pág. 355.

<sup>39</sup> *Ibid.*, pág. 169 y sig.

<sup>40</sup> *Ibid.*, pág. 171. Cfr. al respecto también Nelson, 1988, pág. 151 (n. 3), quien rechaza el término «maximizador restringido» porque opina que o se maximiza o no se maximiza y que, en todo caso, la racionalidad estratégica no lleva a ninguna restricción a la maximización sino, por el contrario, es necesaria justamente para maximizar.

ninguna relevancia porque la racionalidad se refiere exclusivamente a la utilidad (total) de las estrategias, entonces ya no está muy claro en qué sentido justamente la adopción de la estrategia presumiblemente maximizadora de utilidad significa una restricción. Y más aún, si la utilidad de una estrategia se determina, en última instancia, precisamente por el rendimiento (total) que ella proporciona, nos encontramos nuevamente ante el problema empírico de los rendimientos, problema que en cada caso particular hay que resolver mediante un cálculo de los costos y beneficios esperados.

El cambio en la concepción de la racionalidad, es decir, el paso del nivel de la utilidad de los actos particulares al de la utilidad de las estrategias, por cierto, le sirve a Gauthier para inmunizar su argumentación contra cierto tipo de preguntas<sup>41</sup>; pero plantea, a su vez, nuevos problemas fundamentales cuya gravedad hace dudar de la utilidad de estas estrategias para la teoría de Gauthier.

### 3. ¿Es realmente justificatoria la empresa de Gauthier?

Finalmente, aun si la dos preguntas anteriores, acerca de la racionalidad de la maximización restringida, pudieran ser contestadas de manera satisfactoria, quedaría todavía abierta la pregunta más fundamental acerca del *status moral* de la moralidad gauthieriana.

El objetivo de Gauthier es, según sus propias palabras, la «reconstrucción racional» de una moralidad *justificable*, donde «las prácticas morales justificables son aquellas que recibirían nuestro consenso *ex ante*, en una situación premoral adecuada»<sup>42</sup>. En este sentido, se trata, por cierto, de una *empresa justificatoria*<sup>43</sup>.

El resultado de este trabajo justificatorio es una concepción de la moralidad más bien estrecha. La moralidad «justificada» exige tan sólo no aprovecharse de los demás, y cumplir los contratos imparciales, equitativos, para que sea posible la interacción

---

<sup>41</sup> Por ejemplo, resuelve el problema planteado por la aparente contradicción de una disposición de «racionalidad irracional» en el sentido de *Parfit*, 1987, pág. 13.

<sup>42</sup> *Gauthier*, 1988b, pág. 13.

<sup>43</sup> *Gauthier*, 1986, pág. 2; cfr. también *Buchanan*, 1988, pág. 85: según Buchanan, la diferencia básica entre su propio trabajo y el de Gauthier es que el primero es explicatorio, mientras que el segundo es justificatorio.

mutuamente beneficiosa; específicamente, esta moralidad no tiene *nada* que decir acerca de la interacción con aquellas personas que no pueden contribuir nada -o muy poco- al beneficio mutuo.

¿En qué sentido se puede decir, pues, que esta moralidad ha sido *justificada*? El propio Gauthier distingue entre «justificaciones morales» y «justificaciones deliberativas»: las segundas son justificaciones basadas en lo racional; las primeras, justificaciones basadas en lo moral. Gauthier parece, a primera vista, interesarse exclusivamente por el primer tipo de justificación, ya que postula la precedencia de la racionalidad frente a la moralidad: «[L]a justificación deliberativa es claramente más básica... [E]n la medida en que somos agentes racionales, no podemos prescindir de ella, de manera que cuando la justificación moral entra en conflicto con aquélla, la moralidad ... parece oponerse a lo que es racionalmente más fundamental»<sup>44</sup>. Por lo tanto, no puede sorprender que Gauthier considere que se puede justificar una acción o una disposición aduciendo simplemente que ella maximiza la utilidad esperada (basada en las preferencias ponderadas), sin que sea necesaria alguna consideración moral: «La justificación deliberativa ... ignora la moralidad, y aparentemente la sustituye»<sup>45</sup>.

Sin embargo, vista más de cerca, la justificación a la que aspira Gauthier, en última instancia, va más allá de la mera «justificación deliberativa». Gauthier puede sostener que «la justificación deliberativa ... sustituye la moralidad» sólo porque identifica la racionalidad con el criterio último de la justificación *moral*. Ello se muestra muy claramente, por ejemplo, cuando dice: «Evaluar una acción como irracional es, en mi opinión, lo mismo que pretender que ella no debe o no hubiera debido ser ejecutada»<sup>46</sup>. Es precisamente esta concepción normativa de la racionalidad la que, *per definitionem*, crea la coincidencia de lo moral con lo racional y, por consiguiente, también de la justificación deliberativa con la justificación moral.

Las consecuencias de esta concepción de la racionalidad como principio último de la moralidad se manifiestan también cuando Gauthier se plantea la cuestión de «¿qué significa preguntarse a sí mismo cuál es la acción que uno tiene la mejor razón para ejecutar?» y ofrece la siguiente respuesta: «Según mi opinión, las razones tienen que ser derivadas de las preferencias,

---

<sup>44</sup> Gauthier, 1980, pág. 6.

<sup>45</sup> *Ibid.*, págs. 5 y sig., y pág. 8.

<sup>46</sup> Gauthier, 1984, pág. 487.

de manera que, a través de la reflexión acerca de nuestras preferencias, típicamente llegamos a saber cuáles son nuestras razones. *Ceteris paribus*, la mejor razón que tengo es hacer lo que más prefiero. Desde luego, ... una persona puede tener razones para actuar en contra de sus preferencias; insisto en que las restricciones morales se basan en tales razones. Pero una tesis central de mi teoría moral es que estas razones derivan, a su vez, de preferencias; las restricciones a la actividad maximizadora exigen una justificación maximizadora»<sup>47</sup>.

El problema es saber de qué tipo de «razones» Gauthier está hablando; la cuestión se vuelve más grave aún cuando Gauthier afirma, además, que «las razones son causas»<sup>48</sup>. Aquí conviene recordar la distinción propuesta, entre otros, por Carlos S. Nino, entre «razones *explicatorias*» y «razones *justificadoras*»: las primeras son «razones a las que se recurre para explicar o predecir un comportamiento sin valorarlo [y ellas] se superponen con el concepto de motivo», mientras que las segundas son «razones a las que se acude para mostrar la corrección o legitimidad de una conducta»<sup>49</sup>. La afirmación de Gauthier en el sentido de que «las razones son causas» parece indicar que se refiere aquí a razones *explicatorias* o motivos. En este sentido, la llamada «justificación deliberativa» de Gauthier puede ser entendida como una *explicación* basada en el motivo de la maximización de utilidad y, más concretamente, la «empresa justificatoria» de *Morals by Agreement* como una *explicación* de lo que una persona perfectamente racional, es decir, motivada exclusivamente por el motivo de la maximización de utilidad, puede aceptar en términos de restricciones a su actuar.

Ahora bien, las razones *justificadoras* pueden ser causas sólo en la medida en que se convierten, primero, en razones *explicatorias*. Y éstas, a su vez, pueden servir para la *justificación* de una acción sólo si se puede mostrar que forman parte de aquéllas. Pero, al mismo tiempo, una razón explicatoria presupone la *creencia*, por parte de la persona en cuestión, en una razón justificatoria<sup>50</sup>.

Como Gauthier toma en cuenta sólo a las personas perfectamente racionales, puede afirmar *-per definitionem-* que la

---

<sup>47</sup> Gauthier, 1988a, pág. 192.

<sup>48</sup> *Ibid.*, pág. 174.

<sup>49</sup> Nino, 1987, pág. 83. Cfr. con respecto a la confusión de estos dos tipos de razones en Gauthier, Mendola, 1987; sobre la diferencia entre «ser racional» y «tener razones» y las relaciones entre racionalidad, preferencias y valores, cfr. también Baier, 1988.

<sup>50</sup> Cfr. Nino, *loc. cit.*

racionalidad forma parte de sus motivos, es decir, de las razones explicatorias de su comportamiento, lo que implica, además, que estas personas *creen* también en la racionalidad como una -la única- razón *justificatoria*. Por ello, con respecto a estas personas, ciertamente tiene razón Gauthier cuando sostiene que «para establecer un papel explicatorio de la moralidad, hay que demostrar primero sus credenciales justificatorias»<sup>51</sup>, es decir, que no se puede explicar el comportamiento de una persona aduciendo que tenía la intención de comportarse moralmente, sin mostrar que ella cree que el comportamiento moral está también justificado. Pero con ello, volvemos al nivel de la «justificación deliberativa», ya que para las personas racionales de Gauthier, esta es la única «justificación» admisible. Lo que se justifica aquí es, pues, la moralidad; y el criterio justificatorio, la racionalidad.

Sin embargo, lo que queda por mostrar es que -independientemente de lo que crean las personas de Gauthier- la racionalidad tiene, en efecto, la fuerza normativa que le confiere el propio Gauthier; es decir, queda por mostrar que la racionalidad, según los criterios, ya no de la moralidad positiva de las personas, sino de una moral crítica, es una razón justificatoria última. Habría, pues, que justificar la racionalidad, utilizando la moral como criterio justificatorio, es decir, llevar a cabo una justificación moral. Sólo así se podría demostrar que la «moralidad gauthieriana» derivada de la racionalidad merece el nombre de moral. En ningún momento, Gauthier ofrece esta prueba.

Al fin y al cabo, el *status* de la moralidad «justificada» (en el sentido gauthieriano) es más bien lamentable: no puede justificar nada, ya que toda justificación debe basarse sólo en la racionalidad; pero tampoco puede explicar nada, pues obtiene toda su fuerza explicatoria de la misma racionalidad que debe «justificarla». Por lo tanto, la moralidad, así entendida, pierde toda razón de ser.

#### IV. CONCLUSIONES

Resumiendo lo dicho hasta aquí, no es aventurado afirmar que Gauthier no ha logrado mostrar convincentemente ni la racionalidad general de la maximización restringida -es decir, la superación del dilema de los prisioneros en términos puramente racionales- ni la coincidencia de la racionalidad con la moralidad.

*La pregunta del tonto y la respuesta de Gauthier*

73

---

<sup>51</sup> Gauthier, 1988b, pág. 9.



Desde luego, la superación del dilema de los prisioneros aun en ausencia de amenazas externas de sanciones, no depende necesariamente de la corrección general del argumento de la disposición *racional* a cumplir lo acordado. Como recientemente lo ha señalado Robert H. Frank, puede ser una estrategia beneficiosa y eficiente, por ejemplo, convencer a los demás justamente de que uno *no* se deja guiar por la racionalidad, sino por alguna «pasión»<sup>52</sup>.

Por otra parte, para la convivencia con un «Tonto racional» -es decir, para asegurar que a este *siempre* le parezca conveniente, por razones prudenciales, cumplir sus contratos- aparentemente no cabe otro medio que la «solución política», aunque, por razones morales, ciertamente no la propuesta por Hobbes, sino más bien una que parta de la «solución moral» en el sentido de Rawls y asegure el cumplimiento de lo racionalmente acordado, a través de instituciones coactivas también racionalmente acordadas.

En última instancia, el propio Gauthier parece admitir la necesidad de una solución política cuando dice: «En un mundo de Tontos, ser un maximizador restringido y cumplir lo contratado no es beneficioso. En tales circunstancias, ser moral no sería racional»<sup>53</sup>. Como no ha logrado contestar de manera concluyente la pregunta del Tonto, estas son justamente las circunstancias vigentes. Pero, si no es lo mismo ser racional que ser moral, tampoco coinciden la justificación deliberativa y la justificación moral.

Una propuesta que, sin embargo, pretende reconciliar la solución política con la coincidencia de racionalidad y moralidad ha sido presentada recientemente por Otfried Höffe, en su *Justicia Política*. Con respecto al contenido de los contratos racionalmente acordados, Höffe coincide ampliamente con Gauthier; pero, a diferencia de éste, reconoce varios problemas que hacen dudar acerca del cumplimiento espontáneo de lo acordado y, consiguientemente, de la emergencia espontánea de lo que Höffe llama la «justicia natural»<sup>54</sup>. Concretamente, menciona tres argumentos,

---

<sup>52</sup> Frank, 1988.

<sup>53</sup> Gauthier, 1986, pág. 181 y sig.

<sup>54</sup> Señala, sin embargo, que, a primera vista, los problemas no son evidentes (Höffe, 1989, pág. 407): «Por lo pronto, la posición anarquista que sostiene una existencia prepolítica y hasta preinstitucional de la justicia, parece bastante plausible. Pues... debido a sus ventajas para todas las partes, la justicia natural no requiere para su cumplimiento ningún sentido moral, ninguna justicia personal. Puede contentarse con el autointerés como principio motivacional y contar con su imposición espontánea, es decir, preinstitucional.»

de los cuales el más relevante en el presente contexto es el que Höffe llama el «dilema del acatamiento»: «Desde el punto de vista de la teoría de la racionalidad, el dilema muestra que el intercambio mutuo de libertad es óptimo sólo según un cálculo de racionalidad de primer grado, pero, según el cálculo de racionalidad de segundo grado -es decir, según el dilema del polizón- es aún menos que subóptimo, ya que contiene el riesgo de la deserción unilateral ... Allí donde lo único que cuenta es el beneficio, los acuerdos sin medidas coactivas comunes no tienen ningún valor.» La solución propuesta por Höffe es, por lo tanto, una solución política: «Si, con respecto a las obligaciones morales, nos limitamos temáticamente a ... aquellos deberes que surgen de la coexistencia de la libertad, entonces hemos encontrado, en las libertades básicas, una coincidencia de la moral (como justicia) y del autointerés. Esta coincidencia, sin embargo, está puesta en tela de juicio por un cálculo prudencial de segundo grado, es decir, por la ventaja que ofrece la insinceridad o el comportamiento del polizón. Para, a pesar de ello, salvar la coincidencia, se requiere un poder coactivo ... Pero, dado que este poder coactivo es beneficioso para cada cual, su establecimiento ya no es debido sólo en virtud del autointerés, es decir, de un cálculo prudencial de tercer grado; es también una exigencia de la moral (interpretada como justicia)»<sup>55</sup>.

El problema con esta solución de Höffe es que no queda nada claro en qué sentido este poder coactivo es «beneficioso para cada cual» si su función consiste justamente en evitar que alguien pueda beneficiarse del comportamiento de polizón. Como ya se ha visto en el análisis de la propuesta de Gauthier, el que ello sea beneficioso o no para una persona sigue siendo una cuestión empírica, contingente en cada caso particular.

Con respecto a la relación entre racionalidad y moralidad, parece, pues, acertada la afirmación de Harsanyi: «Si nos importa el bien común, entonces la racionalidad nos dice claramente qué código moral hay que seguir ... Pero si no nos importa el bien común, entonces la racionalidad no puede decirnos que hay que seguir éste (ni tampoco cualquier otro) código moral»<sup>56</sup>.

*La pregunta del tonto y la respuesta de Gauthier*

75

---

<sup>55</sup> *Ibid.*, pág. 425 y págs. 425 y sig. Los otros dos problemas con la constitución espontánea de la justicia natural mencionados por Höffe se refieren al surgimiento de «conflictos de interpretación» de las reglas acordadas y al «dilema diacrónico» de la justicia entre las generaciones como así también a la cuestión del tratamiento de aquellas personas que no tienen nada que ofrecer para entrar en una cooperación mutuamente beneficiosa (como, por ejemplo, algunos tipos de minusválidos).

<sup>56</sup> Harsanyi, 1985, pág. 55.

Como he tratado de mostrar, la «solución racional» propuesta por Gauthier para dar respuesta a la pregunta del Tonto conduce a una serie de nuevas preguntas -de las que, desde luego, sólo algunas han podido ser tratadas aquí, aunque creo que ellas son básicas. Por lo tanto, parece que el Tonto, que no deja de tener sus intervalos lúcidos, todavía no ha sido desplazado de su lugar en el panteón mencionado más arriba. Seguirá preguntando, a menos que algún día deje de concentrarse en el valor exclusivo de la racionalidad y adopte para sí mismo un sentido de moralidad.

### BIBLIOGRAFÍA

- Baier, Kurt, 1988, *Rationality, Value and Preference*, en E. Frankel Paul *et al.* (1988), 17-45.
- Blits, Jan H. (1989), «Hobbesian Fear», en *Political Theory*, 17:3 (August), 417-431.
- Braybrooke, David, 1987, «Social Contract Theory's Fanciest Flight», en *Ethics*, 97 (July), 750-764.
- Buchanan, James M., 1988, *The Gauthier Enterprise*, en E. Frankel Paul *et al.* (1988), 75-94.
- Elster, Jon, 1987, *Subversion der Rationalität*, Francfort d. M. y Nueva York.
- Ethics*, 1985, vol. 96, núm. 1, «Symposium on Rationality and Morality».
- Frank, Robert, H., 1988, *Passions within Reason. The Strategic Role of the Emotions*, Nueva York: Norton.
- Frankel Paul, Ellen *et al.* (comps.), 1988, *The New Social Contract-Essays on Gauthier*, Londres y Nueva York.
- Gauthier, David, 1984, «Deterrance, Maximization, and Rationality», en *Ethics*, 94 (April), 474-495.
- , 1986, *Morals by Agreement*, Oxford.
- , 1987, «Taming Leviathan», en *Philosophy, & Public Affairs*, 16:3 (Summer), 280-298.
- , 1988a. *Morality, Rational Choice, and Semantic Representation - A Reply to My Critics*, en E. Frankel Paul *et al.* (1988), 173-221.
- , 1988b. *Why Contractarianism?* Mimeo (del original inglés de la contribución de Gauthier a este número de *Doxa*).
- Hampton, Jean, 1986, *Hobbes and the Social Contract Tradition*, Cambridge *et al.*

- Harsanyi, John C., 1985, «Does Reason Tell Us What Moral Code to Follow and Indeed, to Follow Any Moral Code at All?» en *Ethics*, 96:1 (October), 42-55.
- Hegselmann, Rainer, 1988, *Rational Egoism, Mutual Advantage and Morality - A Review-Discussion of D. Gauthier: Morals by Agreement*, Mimeo.
- Hobbes, Thomas, 1651, *Leviathan*, citado según la edición de C. B. Macpherson, Harmondsworth, Penguin, 1985.
- Höffe, Otfried, 1989, *Politische Gerechtigkeit. Grundlegung einer kritischen Philosophie von Recht und Staat*, Francfort d. M.
- Hume, David, 1739-40, *A Treatise of Human Nature* (edición e introducción de Ernest C. Mossner), Harmondsworth, Penguin, 1987.
- Kliemt, Hartmut, 1987, «The Reason of Rules and the Rule of Reason», en *Crítica*, vol. XIX, núm. 57 (diciembre), 43-86.
- , 1989, *Der Beitrag der Spieltheorie zur Lösung des Hobbesschen Ordnungsproblems*, en Wolfgang Boger (comp.), *Zum Menschenbild der Politischen Ökonomie*, en prensa.
- Kraus, Jody S., y Jules L. Coleman, 1987, «Morality and the Theory of Rational Choice», en *Ethics*, 97 (July), 715-749.
- Lenk, Hans, 1986, «Über Rationalitätstypen und Rationalitätskritik», en (del mismo autor) *Zwischen Wissenschaftstheorie und Sozialwissenschaftstheorie und Sozialwissenschaft*, Francfort d. M., 104-130.
- Locke, John, 1689, *Second Treatise*, citado según la edición de South Bend (Ind.), Regnery/Gateway, 1955.
- McClennen, Edward F., 1989, «Justice and the Problem of Stability», en *Philosophy & Public Affairs*, 18:1 (Winter), 3-30.
- Mendola, Joseph, 1987, «Gauthier's *Morals by Agreement* and Two Kinds of Rationality», en *Ethics*, 97 (July), 765-774.
- Nelson, Alan, 1988, «Economic Rationality and Morality», en *Philosophy & Public Affairs*, 17:2 (Spring), 149-166.
- Nino, Carlos S., 1987, *Introducción a la filosofía de la acción humana*, Buenos Aires.
- Parfit, Derek, 1987, *Reasons and Persons*, Oxford.
- Sen, Amartya, 1982, «Rational Fools: A Critique of the Behavioural Foundations of Economic Theory, en *idem*. Choice, Welfare and Measurement», Oxford, 84-106, primera publicación en *Philosophy & Public Affairs*, 6 (Summer, 1977).
- Taylor, Michael, 1987, *Possibility of Cooperation*, Cambridge *et al.*
- Vanberg, Viktor y James M. Buchanan, 1989, «Interests and Theories in Constitutional Choice», en *Journal of Theoretical Politics*, 1:1 (January), 49-62.

## DOXA -6 (1989)

