

# *Homo Economicus* e Individuo Liberal. Una derivación de la moral a partir del interés propio

## *Homo Economicus and liberal Individual. A derivation of Morals from Self-Interest*

Blanca RODRÍGUEZ LÓPEZ

### **Resumen**

Apoyándose en la Teoría de la Elección Racional y en el contractualismo, Gauthier intenta derivar la moralidad como una restricción en la maximización de la utilidad justificada por la razón a partir de una situación y unos agentes moralmente neutrales. El mantenimiento de los acuerdos satisfactorios sería racionalmente justificable incluso en ausencia de restricciones externas. En este artículo analizamos la argumentación ofrecida por Gauthier en apoyo de su contractualismo moral y defendemos que, en determinadas situaciones, los acuerdos cuyo cumplimiento apoya la razón son, siguiendo a Hobbes, los apoyados por restricciones externas.

*Palabras clave:* Gauthier, Hobbes, Contractualismo, Restricciones morales, Restricciones externas, Racionalidad.

### **Abstract**

From Contractualism and the Theory of Rational Choice, Gauthier attempts to infer morality as a rationally justified restriction of utility maximization when considering a situation involving morally neutral agents. Even in absence of external

constraints, satisfactory agreements should be kept for a purely rational justification. In this paper, we analyze the reasons offered by Gauthier in favor of a moral contractarianism and, following Hobbes, defend that in particular situations the agreements sustained by reason are those that are also sustained by external constraints.

*Keywords:* Gauthier, Hobbes, Contractarianism, Moral constraints, External constraints, Rationality.

## 1. Situaciones DP y Estado de Naturaleza

Supongamos que una noche de invierno, tras una agradable cena con amigos, me dispongo a volver a casa cuando me doy cuenta de que he perdido la cartera. La situación es bastante mala: hace un frío tremendo, llevo un par de copas de más, no conozco a nadie que viva por los alrededores y el móvil se ha quedado sin batería. Mi mejor opción (incluso la única razonablemente buena) es parar un taxi e intentar llegar a un acuerdo. Si me lleva, la pagaré después (tengo dinero en casa) y le daré una buena propina. El acuerdo es también provechoso para el taxista. Con tanto frío no ha tenido muchos clientes y yo le ofrezco la posibilidad de una buena carrera más una propina considerablemente más generosa de lo habitual. Posiblemente muchos nos hayamos encontrado en una situación parecida, con variados desenlaces. Su estructura, relativamente simple, tiene dos características fundamentales:

1. Unos agentes se encuentran en una situación en la que es mejor para todos establecer un acuerdo que no hacerlo.
2. Pueden llegar a un acuerdo en unos términos que todos consideran aceptables y satisfactorios.

Es fácil reconocer en esta estructura un paralelismo con la de dos situaciones que se encuentran entre las más debatidas en la filosofía moral y política, una desde hace siglos y otra desde hace décadas: el Estado de Naturaleza de los filósofos contractualistas y el Dilema del Prisionero de los teóricos de la racionalidad práctica. David Gauthier, sin duda uno de los filósofos contemporáneos más relevantes, reconoce explícitamente este paralelismo afirmando que el Estado de Naturaleza es “un estado con la estructura interna de un Dilema del Prisionero generalizado”<sup>1</sup> y realiza su propuesta filosófica, tan bien conocida como interesante: una teoría contractualista que utiliza los conceptos y supuestos de la Teoría de la Elección Racional (TER).

---

<sup>1</sup> Gauthier “Between Hobbes and Rawls”, p. 26.

Algunos de los supuestos básicos más importantes de la TER giran en torno a la caracterización de los agentes, a los que se supone auto interesados, en el sentido de que tienen objetivos que quieren alcanzar (o, por utilizar términos más característicos de la TER, preferencias que quieren satisfacer) y racionales (capaces de determinar y realizar las acciones más adecuadas para alcanzar sus objetivos). Entendiendo que la Utilidad es una medida que se establece sobre las preferencias de los individuos, la TER define la *acción racional* como aquella que maximiza la utilidad del agente. De forma paralela, los contractualistas clásicos ponen a la base de sus teorías una caracterización de los individuos que habitan el Estado de Naturaleza, siendo la ofrecida por Hobbes en la primera parte de *Leviatan* (“Del Hombre”) la que guarda mayor similitud con la de la TER, y por ende con la de Gauthier.

En estas situaciones, que para abreviar llamaremos *situaciones DP*, la razón que caracteriza a nuestros agentes les lleva a establecer un pacto<sup>2</sup>. Pero lo que saca a nuestros individuos de la desafortunada situación en la que se encuentran no es tanto, o no tan sólo, establecer pactos, sino cumplirlos. Hobbes expresa de forma contundente la necesidad de cumplir los pactos en su tercera Ley de la Naturaleza: “*que los hombres deben cumplir los convenios que han hecho*. Sin esta ley, los convenios se hacen en vano y sólo son palabras vacías”<sup>3</sup>. Esta ley sería para Hobbes el origen de la *Justicia*, ausente en el Estado de Naturaleza, en el cual para cada individuo “su apetito personal es la medida de la bueno y de lo malo”, identificando la justicia con el mantenimiento de los convenios. Paralelamente, Gauthier introduce la moralidad como un medio de restringir la maximización de la utilidad individual ateniéndose a los acuerdos<sup>4</sup>.

Aunque la razón dicta tanto el establecimiento como el cumplimiento de los pactos, Hobbes y Gauthier coinciden en que tal dictamen no se aplica a todos los pactos sino sólo a algunos. Sin embargo, este último punto de acuerdo entre ambos contractualistas marca también el punto de origen de las divergencias, pues las condiciones que deben cumplir los pactos para que la razón exija su cumplimiento no sólo son diferentes, sino que el criterio utilizado pertenece a órdenes distintos. Mientras Hobbes acude a la *seguridad* en el cumplimiento mutuo de los convenios, Gauthier acude al *contenido* de los mismos, y sostiene que el cumplimiento queda exigido por la razón en virtud de este. Para distinguir ambos, llamaremos *acuerdos*

---

<sup>2</sup> Este consejo de la razón queda recogido por Hobbes en las dos primeras Leyes de la Naturaleza (que no son sino las Leyes de la Razón, en el capítulo 14 del *Leviatan*.

<sup>3</sup> Hobbes: op.cit. p. 131.

<sup>4</sup> Por este motivo, plantea su teoría contractualista con Hobbes, y no, por ejemplo, con Locke, como punto de referencia, pues el dibujo del Estado de Naturaleza diseñado por este dista de ser un espacio amoral al postularlo dotado de una Ley Natural y unos Derechos Naturales que proporcionan una guía de conducta a los individuos distinta de la ofertada por sus deseos y fines individuales.

válidos (adjetivo utilizado por Hobbes<sup>5</sup>) a aquellos cuyo cumplimiento exige la razón en virtud de la seguridad de su cumplimiento mutuo y *acuerdos satisfactorios* a aquellos cuyo cumplimiento exige la razón en virtud de su contenido.

## 2. Acuerdos válidos y acuerdos satisfactorios

Antes de analizar la diferencia entre ambas posturas, introduciré una terminología y un esquema de análisis habitual en la discusión contemporánea que, por su simplicidad y claridad, nos ayudarán a centrar el problema. Una vez establecido un acuerdo, cada uno de los agentes implicados (salvo que indiquemos expresamente lo contrario, hablaremos siempre del caso más simple de un acuerdo entre dos agentes), se encuentra con dos opciones: cumplir o no lo acordado. Desde el punto de vista de unos agentes racionales y auto interesados, la opción es relevante en aquellos casos cuya estructura es la de una Situación DP. Tales situaciones requieren para su caracterización completa añadir a las dos anteriores otra característica:

3. Cada uno de los agentes puede obtener un mayor beneficio si no cumple lo acordado mientras el otro sí lo cumple, y el que cumple perdería, a consecuencia del incumplimiento del otro, más de lo que habría perdido en ausencia de acuerdo.

En el ejemplo con el que iniciamos este trabajo, yo obtengo un mayor beneficio de mi deserción unilateral, pues el taxista me lleva a casa y yo me ahorro el dinero prometido, mientras que, si yo hago eso, el taxista está peor que si no hubiera habido acuerdo, pues no sólo no cobra sino que pierde tiempo, esfuerzo y gasolina<sup>6</sup>. Debido a la característica 3, y siguiendo una terminología que ha pasado a ser habitual<sup>7</sup>, llamaremos S (beneficiosa para uno mismo) a la acción realizada en cumplimiento del acuerdo y A (altruista) a la que consiste en su incumplimiento. Sin embargo, dado el estricto paralelismo de la situación de los agentes (ambos autoin-

---

<sup>5</sup> Hobbes: op.cit., capítulo 14.

<sup>6</sup> Pudiera parecer que, en mi ejemplo, el taxista no tiene una opción paralela a la mía puesto que no puede desertar del acuerdo, dado que es el primero en cumplir. Habitualmente, los ejemplos de Situaciones DP se refieren a casos en los que el cumplimiento del acuerdo es o bien simultáneo o bien realizado en condiciones en las que cada uno ignora si el otro ha cumplido/está cumpliendo su parte. El motivo de utilizar mi ejemplo es doble: en primer lugar, tanto Hobbes en (1651) Cap. 14 como Gauthier en repetidas ocasiones utilizan ejemplos similares al mío y, en segundo lugar, tales casos muestran una aplicación de la estructura DP más amplia, que es la utilizada por D. Parfit, de quién tomamos la terminología y el esquema de análisis. Los que no resulten convencidos por estas razones pueden sin duda imaginar alguna posible acción del taxista que le facilitaría no cumplir su parte, una vez que me monto en el taxi, y obtener un beneficio que cumpla la característica tres. En beneficio de quien no tenga mucha imaginación, recuerdo que mi situación incluye soledad, incomunicación y una más que ligera borrachera.

<sup>7</sup> Esta terminología es la utilizada por D.Parfit en "Prudences, Morality and the Prisoner's Dilemma".

teresados, racionales, con interés en establecer un acuerdo y más beneficiados si el otro lo cumple y él no) y suponiendo que ambos son conscientes de dicho paralelismo<sup>8</sup>, al querer ambos maximizar su utilidad haciendo S, se encontrarán con que ninguno cumple lo acordado y la situación a la que llegan es subóptima en sentido de Pareto: puede ser mejor para ambos en otro escenario posible, a saber, aquel en que ambos cumple el acuerdo. Esto es lo que hace que las Situaciones DP sean un dilema: puesto que ambos hacen lo que más le beneficia a cada uno, el resultado será para todos peor de lo que podría ser. Parece por tanto que se cumple lo establecido por Hobbes en su tercera ley: todos están interesados no ya en establecer acuerdo, sino en cumplirlos. Pero, ¿por qué razón nuestros agentes escogerían la opción A en una situación en la que la racionalidad de tal acción está en cuestión, dado que por definición S es la acción que maximiza la utilidad del agente?

Puesto que escoger A supone para el agente restringir su búsqueda de la utilidad individual, es habitual hablar de *restricciones externas* cuando se introduce un cambio en la situación en la que se encuentra el agente, de modo que o bien se hace imposible hacer S o se altera el resultado que el agente puede esperar obtener de S y de *restricciones internas* cuando las preferencias del agente sobre los resultados respectivos de A y S se alteran de forma que prefiere el de A<sup>9</sup>.

Las respuestas de Hobbes y Gauthier a la pregunta por aquellos acuerdos que deben (racionalmente) cumplirse difieren y cada uno selecciona un tipo de solución distintas. Tal y como Gauthier reconoce, Hobbes parece referirse a las primeras como las que hacen posible el mantenimiento de los acuerdos. De hecho, la existencia de restricciones externas es lo que caracteriza a un acuerdo como válido y hace por tanto que su cumplimiento sea racionalmente exigible<sup>10</sup>. Desde el punto de vista teórico, son coherentes con la acción de un sujeto racional, puesto que, al cumplir lo acordado, realiza la acción que maximiza su utilidad. La restricción a su conducta se realiza al establecer un poder coercitivo capaz de imponer sanciones pero una vez establecidas estas, el cumplimiento de los acuerdos ya no es una maximización restringida.

Para Gauthier, la moralidad es un tipo de restricción interna. Lo novedoso de su teoría, y el punto en el que se separa de Hobbes, es considerar que la moralidad así entendida está racionalmente exigida: la razón, que exige la maximización de la utilidad, en las situaciones DP exige una *maximización restringida*. Intenta mostrar que no sólo no hay conflicto entre moralidad e interés, sino que el mantenimiento de la primera sirve al segundo. Esto queda perfectamente expresada al comienzo de

---

<sup>8</sup> Esto recoge los requisitos de racionalidad e información de la Teoría de la elección racional así como la igualdad natural de los hombres en Estado de Naturaleza de los contractualistas clásicos.

<sup>9</sup> En Parfit art.cit. p. aparece un cuadro exhaustivo de las diferentes formas que pueden tomar estas soluciones a los dilemas, que él califica de *políticas* y *psicológicas* respectivamente.

<sup>10</sup> Hobbes op.cit., p.125.

su obra principal “ El deber predomina sobre el provecho, pero la aceptación del deber es realmente provechosa”<sup>11</sup>. La racionalidad, sin embargo no nos exige cumplir todos los acuerdos, sino sólo los que podemos considerar satisfactorios. La cooperación de los individuos genera un beneficio determinado que los participantes pueden acordar repartir de diversos modos. Los individuos por tanto no sólo acuerdan cooperar sino que deben también establecer los términos de la cooperación relativos al posterior reparto de los beneficios. Suponemos que estos términos son objeto de un proceso de negociación. Un acuerdo satisfactorio se caracteriza por su imparcialidad en un doble sentido:

- es el resultado de un procedimiento imparcial de negociación. Gauthier defiende que tal procedimiento sigue el *Principio de Concesión Relativa Minimax*<sup>12</sup>.
- el proceso de negociación debe partir de una situación inicial de regateo imparcial. Tal situación se caracteriza no por lo que los agentes llevan realmente a la mesa de negociación, e.d., por lo que tienen en la situación previa a la cooperación (en el Estado de Naturaleza), sino por lo que tendrían si hubieran respetado la *salvaguarda lockeana*, que restringe la interacción previa eliminando de la misma el uso de la fuerza y el fraude<sup>13</sup>.

Cuando los acuerdos son satisfactorios, la razón nos pide una maximización restringida incluso en ausencia de mecanismos políticos que garanticen el cumplimiento de los acuerdos. Podemos entonces establecer las siguientes definiciones:

- Un *Acuerdo Valido* es aquel cuyo cumplimiento está garantizado por sanciones externas
- Un *Acuerdo Satisfactorio* es aquel cuyo cumplimiento está justificado por sanciones internas.

Esta diferencia da al contractualismo defendido por Gauthier un carácter específico que va más allá del contractualismo político, típicamente ejemplificado por la teoría de Hobbes, hasta lo que podemos considerar un contractualismo moral, “una teoría contractualista capaz de justificar la institución y mantenimiento no sólo de estructuras políticas y legales, sino también de restricciones y, sobre todo, disposiciones (modos de elección) personales que pueden denominarse “morales” ”<sup>14</sup>. El Principio de Concesión Relativa Minimax y la salvaguarda lockeana han sido ampliamente debatidos, tanto en lo referente a su carácter de condiciones racionalmente exigibles a un acuerdo cooperativo como su pretendida neutralidad moral<sup>15</sup>.

<sup>11</sup> Gauthier *Morals by agreement*, p.2.

<sup>12</sup> Para una discusión detallada de este principio y los argumentos que lo apoyan, cf. Capítulo V.

<sup>13</sup> En relación a este punto, cf. Gauthier Capítulo VII.

<sup>14</sup> Francés (ed) *Egoísmo, moralidad y sociedad liberal*, p. 18.

<sup>15</sup> Lo más relevante de este debate, con respuestas del propio Gauthier, se encuentra recogido en Gauthier & Sugden (eds), *Rationality, Justice and the Social Contract*.

En este trabajo aceptaremos ambos (como racionalmente exigibles y moralmente neutros) y nos centraremos en discutir si un agente racional debe, en ausencia de restricciones externas, cumplir los acuerdos satisfactorios convirtiéndose en un maximizador restringido.

### 3. Cooperación y beneficio

Podría deducirse de lo dicho hasta aquí que Gauthier defiende que deben cumplirse los acuerdos satisfactorios pese a que sean, en términos de Hobbes, inválidos. Esto sería sin embargo una conclusión apresurada. Para ambos contractualistas, los acuerdos sólo deben cumplirse cuando los individuos tienen una garantía (si no plena, al menos razonable) de que los demás van a cumplir su parte. La diferencia estriba en que, para Hobbes, esta garantía, que convierte los acuerdos en válidos, pasa por el establecimiento de unas medidas de tipo político, mientras que para Gauthier el carácter satisfactorio de los acuerdos es suficiente garantía de cumplimiento mutuo, suponiendo que los agentes sean, tal y como estamos suponiendo, racionales. No obstante, hemos visto que en las situaciones DP, una vez establecido un acuerdo, y por muy satisfactorio que este sea, cada uno de los agentes se beneficiaría de un incumplimiento unilateral. Esta circunstancia, bien vista por Hobbes, inclina a este al establecimiento de restricciones externas porque “el que cumple primero no tiene garantía de que el otro cumplirá después, ya que los compromisos que se hacen con palabras son demasiado débiles (...) si estos no tiene miedo a una fuerza superior con poder coercitivo”<sup>16</sup>. Del Estado de Naturaleza sólo se sale estableciendo una sociedad civil con un poder político suficiente para garantizar el cumplimiento de los acuerdos.

Para defender su postura, Gauthier intenta encontrar un motivo auto-interesado para respetar los acuerdos satisfactorios. El mantenimiento de los acuerdos supone una maximización restringida, pero, por muy restringida que sea, para que pueda defenderse desde el punto de vista de la racionalidad ha de seguir siendo maximización. El intento de Gauthier sigue los siguientes pasos:

1. Un agente racional respetará los acuerdos ya que de otro modo será excluido de acuerdos posteriores. Gauthier no niega que sea beneficioso utilizar la estrategia no cooperativa y realizar la acción S. Lo que afirma es que, a pesar de la ventaja que pueda suponer esto en un caso aislado, a largo plazo esta conducta resulta ser poco rentable, ya que un individuo que haya roto un acuerdo será castigado por los demás con la desconfianza y la consecuente exclusión de acuerdos cooperativos, de modo que acabará perdiendo más de lo que gana.

---

<sup>16</sup> Hobbes op. cit., p. 125.

Sin embargo, esta argumentación solo apoya la racionalidad del mantenimiento de los acuerdos (tanto satisfactorios como insatisfactorios) en un tipo de circunstancias. En efecto, un individuo que incumple un trato será excluido de otros acuerdos, lo cual sin duda le hará perder más de lo que gana. Pero esto sólo sucederá a) si el individuo en cuestión es descubierto, cosa altamente improbable en numerosos dilemas multipersonales b) si es posible aplicar la sanción, e.d., si hay una probabilidad suficientemente alta de que los mismos individuos vuelvan a interactuar y c) si el número de interacciones entre los mismos individuos es indeterminado y desconocido por todos los participantes. Con esto no negamos que la cooperación, entendida como conducta cooperativa, no pueda surgir y hacerse estable en algunas circunstancias en ausencia de restricciones externas<sup>17</sup>. Lo que negamos, es que esto suceda siempre. Más bien, sucede en muy raras ocasiones.

Para hablar de estas situaciones, los teóricos hablan de *Dilemas del Prisionero Iterativos* y hay acuerdo en que, en tales casos, siempre es racional realizar A. Pero el propósito de Gauthier es más ambicioso y pretende abarcar todas las situaciones DP. “ Mi argumento es aplicable a un conflicto que ocurra una sola vez”<sup>18</sup>

2. Un agente racional respetará los acuerdos si a) su disposición a cumplirlos es condición necesaria para que los demás cumplan su parte (es decir, si los agentes son *cooperadores condicionales*, tal y como lo exige la racionalidad, que en ningún caso promueve una cooperación unilateral, cuyo resultado sería para el cooperador incondicional el peor de los posibles) y b) esta disposición puede ser fácilmente descubierta por los demás. Si una persona que se encontrara en una situación DP “pudiera identificar con certeza la disposición de los demás agentes en esa situación, estaría racionalmente justificado para ella adoptar la disposición condicionalmente cooperativa”<sup>19</sup>. Este segundo paso de la argumentación muestra una nueva restricción de las situaciones en las que es racional cumplir los acuerdos. Tal conducta es racional en situaciones DP no iterativas siempre y cuando los agentes sean *transparentes*, e.d., no puedan ocultar a los demás sus disposiciones. Tal situación ideal es, por ello mismo, irreal y Gauthier desplaza el argumento de la transparencia a la *translucidez*<sup>20</sup>. Ser translúcido significa que, aunque los demás no puedan conocer con certeza tus disposiciones, estas tampoco pueden ocultarse por completo (seríamos entonces *opacos*), sino que pueden, de algún modo, intuirlos. Sin embargo, Gauthier ni aclara mediante qué mecanismos pueden ser intuidas nuestras disposiciones ni muestra que de algún modo lo hagamos. Y no parece que podamos

---

<sup>17</sup> Ejemplos de este surgimiento espontáneo de la cooperación y un estudio detenido de las condiciones que posibilitan este surgimiento pueden encontrarse en Axelrod *La evolución de la cooperación*.

<sup>18</sup> Gauthier “El egoísta incompleto”, p. 108.

<sup>19</sup> Gauthier *Ibid.*

<sup>20</sup> Gauthier *op.cit.*, p. 174.

afirmar que los agentes racionales reales sean translúcidos. Lo único que podemos afirmar es que, dado el beneficio que puede esperarse de la cooperación, un agente racional intentará “desarrollar su pericia” para adivinar las disposiciones de los demás y dar a conocer las propias. Pero aunque es fácil defender la racionalidad de lo primero, no sucede así con lo segundo. Mas bien, deberíamos decir que, dado lo beneficioso del incumplimiento unilateral, el agente racional tratará de adivinar las disposiciones de los demás y al tiempo ocultar las propias<sup>21</sup>.

3. El tercer paso dado por Gauthier presenta algunas novedades respecto a los anteriores. En lo que resta del artículo, nos centramos en este, por ser el menos debatido y no, desde luego, el menos interesante. Gauthier es consciente de que si se considera la moralidad como un mero instrumento para el fin del interés propio, entonces se considerará como un mal necesario, un constreñimiento de la búsqueda del propio beneficio que hay que aceptar para evitar males mayores. Pero si esto fuera así, entonces la conducta moral sólo resultaría atractiva para aquellos que no pueden imponer sus propios términos y para los que saben que serán descubiertos y castigados si actúan de modo inmoral. Para evitar esta consecuencia poco deseable, Gauthier propone una nueva visión del ser humano, distinta de la presentada en la figura del hombre económico y, según él, más cercana a la realidad.

Para un ser asocial que busca y lucha, la moralidad no puede ser más que un constreñimiento necesario pero que no es bien recibido. Pero para los que valoran la participación la moralidad del acuerdo, a pesar de ser un fuente de constreñimiento, hace que su actividad compartida sea mutuamente bienvenida y, por tanto, estable, asegurando de este modo la ausencia de coerción y engaño.<sup>22</sup>

El texto parece sugerir que el *individuo liberal* (nombre que se da a este tipo de individuo como figura contrapuesta al hombre económico) valora la participación por sí misma, lo cual hace que su función de utilidad se altere de tal modo que la conducta no cooperativa deja de resultar ventajosa. Se trata por tanto de una sanción interna. Veamos paso a paso como se llega a esta afirmación.

#### 4. El individuo liberal

La argumentación de Gauthier en su defensa del individuo liberal, presentada en el último capítulo de *La moral por acuerdo* es compleja y puede analizarse del siguiente modo.

<sup>21</sup> El propio Gauthier, a pesar de su defensa de la translucidez como una característica a cultivar por los agentes racionales, se muestra consciente de la debilidad de su argumento en este punto (Cf. Gauthier “El egoísta incompleto”).

<sup>22</sup> Gauthier op. cit. p.337. La cursiva es mía.

**A.** Parte de la argumentación consiste en una afirmación de hecho, a saber, que los seres humanos en muchas ocasiones encuentran satisfacción en la cooperación.

En muchos y diversos casos, desde las actividades deportivas a las musicales, desde las políticas a las militares, los seres humanos encuentran satisfacción en la participación. Desde luego, la participación debe estar relacionada con algún fin, pero sería erróneo pensar que la participación tiene un valor meramente instrumental. De hecho, alguien puede valorar los fines en parte porque proporcionan ocasión para la participación.<sup>23</sup>

La cuestión fundamental es si esta valoración de la participación es una solución a las situaciones DP. Para saber esto, antes es preciso saber si la participación tiene valor como medio o como fin.

Es necesario distinguir aquí dos cuestiones distintas: a) si la participación es siempre participación para algo y b) si la participación se valora como fin o como medio. Según se afirma en el texto, la contestación a la primera pregunta es afirmativa. Sin duda lo es. Uno participa con los demás en un equipo de fútbol para jugar al fútbol, una bailarina colabora con un bailarín para bailar un paso a dos. No se puede participar y no hacer nada participando. Lo que se hace a través de la participación es su producto. En ocasiones podemos hablar de este producto como del *fin* de la participación, siempre y cuando no asociemos la palabra “fin” con algo que es valioso por sí mismo.

Esto nos lleva a la segunda cuestión. En ocasiones, la cooperación sólo tiene un valor como medio, e.d., sólo es valiosa en tanto que sirve como medio para conseguir un fin. Pero en otras ocasiones, no sólo la participación es el fin, esta vez entendido en el sentido de ser lo valioso por sí mismo, sino que el producto de la participación es casi una excusa, algo que sólo tiene valor en tanto que da lugar a la participación y, por consiguiente, algo que tiene valor de medio. El mismo caso de tarea colectiva puede servir para ejemplificar ambos tipos de situaciones. Así, la participación con otros en un partido de fútbol puede ser valorada como un medio para lo que aparece como valioso en sí mismo: jugar un partido. Pero también puede ser que jugar el partido sólo sea una ocasión para conseguir algo que se valora por sí mismo: participar con otras personas en algo. Lo importante es que las dos cuestiones son distintas e independientes y no debe cometerse el error de confundir la evidente respuesta afirmativa a la primera como una respuesta afirmativa a la segunda. Una cosa es que lógicamente el hecho de participar requiera un complemento del tipo para qué o en qué y otra distinta es cuál de las dos cosas consideramos que tiene valor en sí misma y cuál tiene un valor derivado, o, dicho con la terminología habitual, qué tiene valor de medio y qué tiene valor de fin.

Hasta el momento, lo único que ha argumentado Gauthier es que la parti-

---

<sup>23</sup> Gauthier op. cit., p.325.

participación no siempre se valora como medio. Esto es cierto. Pero esto, junto con la afirmación de que de hecho la gente en ocasiones valora la participación como fin, al igual que cualquier otra afirmación de hecho respecto a las preferencias de los individuos, no resuelve nuestro problema. Cuando en una situación de interacción lo valioso es la cooperación misma, el agente seguirá la estrategia cooperativa. Pero entonces ya no hay motivo para llamar a esta estrategia A(ltruista), pues al seguirla el agente no obra de un modo beneficioso para los otros jugadores, sino para él mismo (puede que esto también, y de modo lateral, beneficie a los demás jugadores, pero este no es el motivo por el que un jugador escoge esa estrategia). Una situación en la que todos ganan con la cooperación y ninguno tiene ningún incentivo para obrar de forma no cooperativa no es una situación DP. Desear la cooperación como fin no soluciona el dilema, sencillamente porque no hay ningún dilema que resolver.

**B.** La argumentación de Gauthier no se reduce a la anterior afirmación de hecho. Más bien lo que intenta es explicar por qué un agente racional auto-interesado para quien la cooperación sólo tiene un valor instrumental (cosa que incluso puede muy bien sucederle al individuo liberal) ha de cooperar y ha de encontrar valiosa la cooperación. De nuevo, valorar la cooperación se entiende como una característica moral que debe ser fundada en el interés. Y, desde luego, si se argumentara que la conducta moral se basa en el interés porque tenemos interés en mantener una conducta moral, el argumento sería vicioso<sup>24</sup>.

¿Por qué entonces debe un individuo racional valorar la cooperación y la participación en tareas colectivas? Según Gauthier, una persona valora la cooperación (debemos entender que como medio) sólo si considera que esta es un *costo necesario* para conseguir algún fin valioso. Ese fin valioso por sí mismo es la maximización de la utilidad. Pero, como hemos visto, un agente maximiza su utilidad si no coopera, puesto que la estrategia no cooperativa es mejor para el agente bajo cualquier supuesto acerca de la conducta de los demás. Sin embargo, la conducta no cooperativa generalizada es peor para todos. Este puede ser el motivo por el que se afirma que la cooperación es necesaria para conseguir un fin valioso. Cada uno de los agentes sabe que, supuesto que todos son agentes racionales, el resultado obtenido será el no cooperativo, un resultado subóptimo. En este sentido, la cooperación es valiosa. “Por supuesto”, pensará cada agente individual, “sobre todo la cooperación de los demás. Si los demás cooperan y yo no, el resultado será para mí el mejor de los posibles. Disfrutaré de los beneficios sin tener que pagar los costos”. Pero, de nuevo, cada uno sabe que los demás piensan exactamente lo mismo y que, por

---

<sup>24</sup> Debe notarse que la solución 3 no da lugar necesariamente a un argumento circular. Lo único que se dice es que puede suceder que valoremos la conducta moral por sí misma de tal modo que al obrar moralmente obramos en nuestro propio interés. Cuando se da esta situación, moral e interés coinciden.

tanto, la cooperación, cuyo carácter valioso no se duda, no se conseguirá. ¿Qué puede hacer un agente racional?. Sin duda, puede intentar establecer sanciones externas o desarrollar una disposición que suponga una sanción interna. Gauthier opta por esta última y parece afirmar que, dado lo valioso de la cooperación, esta actitud se desarrollará espontáneamente. Pero ¿cómo y por qué?

C. “Ciertamente las actividades que tienen un valor instrumental deben tener un sentido más allá de ellas mismas, y este sentido no puede ser meramente el compromiso con estas actividades. Debemos por tanto reconocer ciertos estados de cosas intrínsecamente valiosos, deseados por sí mismos. Pero no tenemos que suponer que esos estados de cosas, por muy valiosos que sean, son suficientes para llenar una vida que los seres humanos encuentren digna de ser vivida.”<sup>25</sup>

Este argumento no muestra que un agente racional que sólo se interesa en la cooperación como medio deba cooperar. Más bien la argumentación está dirigida a mostrar que, puesto que la vida humana es más valiosa si los hombres cooperan, un individuo racional debe valorar la cooperación. Hay un sentido en el que no puede dudarse de la verdad de esta afirmación. Si valoramos la cooperación y cooperamos, la satisfacción que la vida puede proporcionarnos aumenta. Esta argumentación es similar a la que a veces se les hace a las personas a las cuales no les gusta el vino: “No sabes lo que te pierdes. Si te gustara el vino, esto te daría ocasión de aumentar la cantidad de satisfacción en tu vida. Disfrutarías, además de los placeres de los que disfrutas ahora, del placer del vino”. Esto probablemente es cierto, al menos si el gusto por el vino es lo suficiente para disfrutar de él cuando llega la ocasión y no tan fuerte como para padecer por su ausencia. Tomado en este sentido, la exhortación a que gustemos del vino, la cooperación o la ópera, es un buen consejo.

Sin embargo, hay otro sentido en el que esto es falso. Puede darse una lectura del texto según la cual se afirma que debe valorarse la cooperación en tanto que es indispensable para conseguir algún fin valioso. El problema está en qué debemos entender aquí por “indispensable”. En efecto, con esta expresión podemos referirnos a dos cosas distintas:

1. Pensemos en nuestro ejemplo inicial. El resultado (llegar a casa sana y salva, “arreglar la noche” con una buena carrera”) es valioso en sí y sólo puede ser conseguido mediante la cooperación. Pero la cooperación no tiene nada que ver con resultado. De hecho, el mismo resultado podría conseguirse de otro modo (un amigo que vuelve, otro cliente). El resultado en sí es independiente de la cooperación. La conexión que une el medio (la cooperación) al fin (el resultado deseado) no es de necesidad. En otros casos esta conexión es de lo que habitualmente se conoce como “necesidad física”<sup>26</sup>. La conexión de la cooperación con la consecución del fin pare-

<sup>25</sup> Gauthier op. cit., p.332

<sup>26</sup> Esto sucede, por ejemplo, en el siguiente caso. Imaginemos un grupo de personas que quiere coger algo escondido en un pozo a 25 metros de profundidad. Si todas colaboran en la tarea, el botín

ce en estos casos bastante sólida. De hecho lo es. Pero ambas cosas, medio y fin, siguen siendo lógicamente independientes. Llamaremos a este tipo de actividades *Actividades cooperativas*

2. Hay, sin embargo, otro tipo de casos en los que esta conexión es de necesidad lógica o necesidad en sentido fuerte. Supongamos que yo quiero bailar un paso a dos. Bailar un paso a dos significa bailar con otra persona. Es imposible que una sola persona baile un paso a dos. Esta imposibilidad no viene sólo dada por el hecho de que, con la ayuda de otra persona, yo pueda hacer cosas que no puedo hacer sola, como por ejemplo realizar ciertos giros. Quizá alguien pueda hacerlo o alguien algún día puede que lo haga. Pero estos resultados deseados, igual que en los casos anteriores, sólo de un modo que podríamos llamar accidental tienen algo que ver con la cooperación con otra persona. La imposibilidad de bailar sólo un paso a dos es una imposibilidad lógica. Hay determinado tipo de actividades que requieren la cooperación de un modo esencial. Las llamaremos *Actividades participativas*

La conexión entre cooperación y resultado en las actividades cooperativas puede describirse como una relación de medios a fines. En estos casos, la cooperación es sólo un medio. Puede que, por el motivo que sea, ese medio sea de hecho necesario, pero al ser necesario sólo en tanto que medio, la necesidad de su conexión con el fin no elimina la independencia lógica de ambos. Sin embargo, en las actividades participativas, hablar de medios y fines pierde su sentido. Si yo quiero bailar un paso a dos, entonces tengo que bailar con otra persona. Pero bailar con otra persona no es un medio para bailar un paso a dos. Más bien, bailar con otra persona (ateniéndose a determinadas reglas) es bailar un paso a dos. La relación entre ambas cosas no es causal, sino formal. En realidad, esto es lo que quiere decir que la relación entre dos cosas es de necesidad lógica.

Recordemos que lo que nos interesa es saber si tiene razón Gauthier al afirmar que un agente racional valorará en sí la cooperación cuando esta es un medio para conseguir algún fin valioso por sí mismo. La distinción entre estos dos tipos de casos es sumamente importante para resolver nuestra cuestión<sup>27</sup>. Porque la relación de preferencia se establece de un modo muy distinto en uno y en otro. Kant afirmaba que el que quiere un fin quiere también los medios indispensables para alcanzarlo<sup>28</sup>. Kant se refiere a aquellos medios cuya relación con el fin se expresa median-

---

debe ser repartido. Cada uno preferiría quedarse con todo, pero el problema es que el tesoro sólo puede ser conseguido mediante la cooperación, e.d., si forman una cadena agarrándose unos a otros (supongamos para simplificar que la situación hace imposible el uso de escalas etc). Uno sólo podría conseguirlo si algo cambiara en la naturaleza, por ejemplo si uno de ellos midiera al menos 24 metros, si pudiera dar un salto vertical de esa longitud o si pudiera volar.

<sup>27</sup> No debe confundirse esta distinción con la establecida más arriba. No se trata ahora de que la cooperación pueda ser en ocasiones un fin en sí misma. Más bien se trata de que hay casos en los que no hay una relación causal de medios a fines entre la cooperación y algo que sea resultado de la misma ni en una dirección ni en otra.

<sup>28</sup> Kant *Fundamentación de la metafísica de las costumbres*, p.66.

te proposiciones sintéticas.

Que para dividir una línea en dos parte iguales, según un principio seguro, tengo que trazar desde sus extremos dos arcos de círculo, es cosa que la matemática enseña, sin duda por proposiciones sintéticas; pero una vez que sé que sólo mediante esa acción puede producirse el citado efecto, si quiero íntegro el efecto, quiero también la acción que es necesaria para él, y esto último sí que es una proposición analítica, pues es lo mismo representarme algo como efecto posible de cierta manera por mí y representarme a mí mismo como obrando de esa manera con respecto al tal efecto.<sup>29</sup>

Tales proposiciones sintéticas expresan auténticas relaciones causales. Es más, ninguna relación auténticamente causal puede expresarse mediante proposiciones analíticas. El primer problema que surge respecto a estas relaciones causales es que es sumamente difícil decir si son o no necesarias y en qué sentido. Esto sucede porque, en tanto que no haya ninguna conexión esencial entre medio y fin, siempre existe la posibilidad de obtener el fin utilizando otro medio. Sea como sea, hay muchos casos en los que la relación entre un medio y un fin (entendidos ambos en sentido propio) es necesaria de hecho, e.d., en los que la consecución del fin en un momento determinado pasa imprescindiblemente por la utilización de determinados medios. Supongamos que esto es suficiente para decir que un medio es “indispensablemente necesario para alcanzar un fin”.

¿En qué sentido puede decirse que, en estos casos, el que quiere el fin quiere (o, mejor, debe querer) los medios? Hay un sentido trivial en el que esto es verdad: el que quiere el fin tiene que estar dispuesto a poner los medios. Pero hay otro sentido en el que esto no es cierto. En este otro sentido, no sólo el que quiere el fin no tiene por qué querer los medios, sino que incluso, y en muchas ocasiones, quiere el fin a pesar de los medios que es necesario disponer, e.d., a pesar de no querer los medios. Esto es precisamente lo que diferencia querer algo como fin y quererlo como medio. Lo que se quiere como medio se quiere en el sentido de que se está dispuesto a realizarlo, pero no en el sentido de que su utilización sea, digamos, deseada. En muchos casos el empleo de determinados medios supone un coste a descontar del total de utilidad que supone la consecución del fin. Por eso, cuando hay más de un medio disponible, se escoge aquel que resulta menos costoso. Cuando sólo hay uno, no existe esta posibilidad de elección, pero el medio puede seguir resultando costoso. Esto sucede en todos los casos en los que el medio tiene sólo valor como tal. Como mucho, en algún caso puede no resultar costoso, sin que esto suponga que tiene un valor positivo.

En los casos de actividades participativas, en los que la relación no es causal sino formal ocurre otra cosa muy distinta. Pero no porque el que quiera una cosa

---

<sup>29</sup> Kant *Ibid.*

deba querer la otra, sino porque el que quiere una cosa, de hecho, está queriendo la otra. Si yo quiero bailar un paso a dos entonces quiero bailar con otra persona. Si quiero jugar un partido de fútbol, entonces quiero jugar con otras personas. Pero aquí no hay fines ni medios en sentido propio. No hay relación causal.

Sin duda, y a juzgar por los ejemplos utilizados por Gauthier, cuando este afirma que, al querer un fin valioso por sí mismo, un agente racional querrá también los medios, esta pensando en este tipo de casos. Uno de sus ejemplos más utilizados es el de la orquesta, donde afirma que “incluso si un sólo individuo pudiera llegar a dominar todos los instrumentos, no podría convertirse en una orquesta”<sup>30</sup>. Gauthier habla como si en estos casos la cooperación fuera un medio necesario. Pero no lo es. No es un medio de ningún tipo. Más bien el fin que se quiere es un fin participativo. Yo quiero tocar en una orquesta. Esto mismo ya es querer participar con otros<sup>31</sup>. Pero no es que yo valore en sí un medio. Gauthier afirma que el individuo liberal, en estos casos valorará el medio, la participación. Esto es una de las características que le distinguen de otro tipo de hombre, el hombre económico. Pero, por todo lo dicho, esto no es así. Naturalmente que el individuo liberal valora la participación en estos casos, pero esto no le distingue de los demás. Por el contrario, todo aquel que se proponga como fin una actividad participativa lo hace. Porque en estos casos, la afirmación no es, como él pretende, una afirmación de carácter psicológico. La psicología no tiene nada que ver aquí. Es una afirmación de tipo lógico, en la que se muestra cuál es la implicación lógica de querer un fin colectivo, a saber, el querer lo que, o bien forma parte de este fin, o bien es el fin mismo.

**D.** El siguiente paso en la argumentación consiste en afirmar no sólo que cuando un individuo quiere uno de estos fines quiere por ello la participación, sino que estos fines son tales que deben ser queridos.

¿No podemos más bien suponer que nuestra capacidad social para encontrar valor en la participación es una de las principales fuentes de enriquecimiento en la vida humana, haciendo posible, como lo hace, la realización complementaria de nuestras variadas capacidades y potencias humanas?<sup>32</sup>

Naturalmente, la contestación de Gauthier es afirmativa. Es importante notar que el planteamiento de la cuestión retrotrae la argumentación al punto que discu-

---

<sup>30</sup> Gauthier op. cit., p.336.

<sup>31</sup> Es notable que, al hablar de estos casos, Gauthier utilice indistintamente los términos “cooperación” y “participación”. Sin embargo, es importante notar que no son equivalentes. Tanto los casos que plantea Gauthier como los que yo presento son casos de participación. Pero esto no excluye la posibilidad de que se planteen situaciones en los que ambos conceptos difieran. Puedo querer jugar una partida de póquer. Esto significa que quiero participar. Pero también puedo utilizar con los demás una estrategia no cooperativa, como hacer trampas.

<sup>32</sup> Gauthier op. cit., p.337.

timos más arriba, cuando se planteaba si debe valorarse la cooperación. La respuesta, en este caso, vuelve a ser la misma: al igual que adquirir gusto por el vino o la música clásica o cualquier otra ampliación del repertorio de las cosas capaces de producirnos satisfacción, adquirir gusto por las tareas colectivas incrementa nuestras posibilidades de satisfacción. Esto es cierto. Incluso podemos dar por bueno que estas tareas no sólo son una fuente de enriquecimiento, sino que es una de las principales, siempre y cuando se entienda correctamente. Es una de las fundamentales fuentes de enriquecimiento porque es una de las más numerosas y variadas. Pero la cuestión acerca de si las satisfacciones de este tipo son “en sí” más valiosas o no lo son es mucho más discutible. De hecho, la TER afirma que las preferencias (racionales) de los agentes sólo son más o menos valiosa en tanto que estos las colocan en uno u otro lugar en la ordenación. Pero esto hace aun más dudosa la afirmación de que un agente racional debe valorar estas actividades, igual de dudosa que la afirmación de que un agente racional debe valorar el vino. Esto es cuestión del contenido de las preferencias, y la racionalidad, al menos tal y como la presenta la TER, y que Gauthier acepta, no hace ningún supuesto acerca del contenido de las preferencias. No resulta lícito hacerlo ahora sólo para salir del paso.

Ateniéndonos a los casos de actividades cooperativas, ¿por qué no considerar la conducta cooperativa como un mal menor, como un medio cuyo necesario empleo puede restar valor al resultado? Según Gauthier, esto sucede porque el resultado de la cooperación se considera únicamente como preferible en segundo lugar. Pero, añade, esto es falso. De hecho, pretende mostrar que el resultado de la conducta cooperativa es el mejor resultado. El argumento fundamental es que la cooperación nos permite conseguir resultados que, sin ella, estarían fuera del alcance de las posibilidades humanas. Y este hecho no es de lamentar, pues nos permite gozar de la cooperación. Pero aunque es cierto que el resultado de la cooperación no puede ser alcanzado sin esta, que esto no sea lamentable depende de si se desea la cooperación con independencia del resultado. Y, en caso contrario, el resultado de la cooperación sí es lo que prefiero en segundo lugar. Pensemos en mi ejemplo del taxista. Ninguno desea cooperar (yo prefiero un príncipe azul y él encontrarse una billetera). Y cooperar supone un coste. Por eso, el resultado de la cooperación es la segunda mejor opción. Para cada uno hay un resultado mejor que el resultado cooperativo, aquel que se sigue si el otro cumple lo acordado y uno no. Esto sucede en todas las situaciones DP. Naturalmente, hay situaciones que no son de dilema. En ellas el resultado “cooperativo” es el mejor, pero lo que necesitamos es una teoría que solucione los dilemas, no que suponga que estos no se dan.

**E.** La argumentación de Gauthier fracasa en demostrar precisamente lo que era su tesis principal, a saber, que cuando la participación se valora como medio, pasa (o, incluso, debe pasar) a valorarse como fin. Cuando la actividad es cooperativa la cooperación desempeña el papel de medio, y entonces no hay ningún motivo para

que pase a valorarse como fin. Sin embargo, el argumento no se detiene en este punto. Porque, según él, de la valoración de la participación, e.d., de la valoración que no es meramente instrumental, se pasa a valorar a los que participan con nosotros.

Pero además de esto, podemos suponer que, al valorar la participación, una persona llega a valorar a sus compañeros, de tal modo que las actividades compartidas dan lugar a vínculos entre las personas que conducen a cada uno a tomarse interés en los intereses de los otros- aunque, desde luego, no en el interés de todos los otros sino sólo de aquellos a los que se identifica y experimenta como co-participantes.<sup>33</sup>

Esto sucede porque

Las personas llegan a tomarse interés en sus compañeros porque reconocen la mutua disposición a no aprovecharse unos de otros y a compartir de un modo justo los beneficios conseguidos mediante una actividad compartida.<sup>34</sup>

Los agentes escogerían A debido a que ahora son lo suficientemente altruistas y, por ello, A ya no es peor para ellos. Como los intereses de los demás llegan a valorarse a partir de la valoración de la cooperación, y esta es racional, podemos decir que la moral se basa en el interés propio o, dicho de otro modo, podemos afirmar que partiendo del supuesto de unos agentes auto-interesados hemos llegado a derivar la moralidad. Lamentablemente, la argumentación no parece concluyente.

Para Gauthier, el hecho de participar en tareas conjuntas con otras personas, junto con el hecho de que estas se muestren dispuestas a basar la cooperación en unos términos justos, es suficiente para que cada uno desarrolle un interés altruista en los intereses de los demás. Pensemos en el paso a dos. El “objetivo” de la cooperación es bailar un paso a dos. Pero para bailar un paso a dos hay que contar con otra persona, por tanto, puesto que ambos quieren bailar un paso a dos, ambos quieren cooperar. Si los dos cooperan, se baila el paso a dos, y si alguno o ambos no coopera, no se baila (suponemos para simplificar que sólo hay dos bailarines disponibles). Pero si ambos quieren bailar, entonces no nos encontramos en una situación DP al faltar la tercera característica de estas.<sup>35</sup> Lo mismo sucede en los casos

---

<sup>33</sup> Gauthier op. cit., p.336. La cursiva es mía.

<sup>34</sup> Gauthier op. cit., p.338

<sup>35</sup> Naturalmente, dos personas en esta situación pueden tener algunos intereses contrapuestos, como por ejemplo el puesto en el que aparecerán en el cartel, el tipo de letra en que irá su nombre o como saludarán. Si ponerse de acuerdo en estas cuestiones depende de ellos (cosa que no suele ocurrir) deberán entablar un proceso de regateo. El resultado será garantizado en este caso por la existencia de un contrato. Pero el regateo no tendrá como objeto el determinar si bailan o no ni hasta que punto bailan. Más bien, es probable que el bailar o no sea utilizado por cada uno como una amenaza

en los que la propia cooperación es el fin deseado por todos los jugadores<sup>36</sup>.

Por consiguiente, en estos casos el dilema no se plantearía. El hecho de que cada jugador se interese o no en el bienestar de los otros no soluciona ningún problema porque no hay ningún problema que resolver. La afirmación de Gauthier es en estos casos superflua. Pero, además, la generalización del caso de las actividades participativas y del caso en el que la cooperación es el resultado deseado a la totalidad de las situaciones en las que se plantea la cooperación es ilegítima, al tomar toda su plausibilidad de casos en los que la propia cooperación es (al menos parte fundamental) del objetivo de la misma. Por una razón. Cuando para alguien el producto de la cooperación es sólo una excusa para la cooperación misma, lo más habitual es que esa persona sienta una cierta *simpatía previa* por las personas con las que quiere colaborar. Pero no es que la cooperación haga surgir la simpatía. No es una *simpatía inducida* por la cooperación. De modo que la existencia del interés por los demás en casos de simpatía previa no prueba lo que Gauthier quiere probar, a saber, que es la actividad compartida la que genera la simpatía. Y, en el resto de los casos, incluidos la mayoría de las actividades participativas<sup>37</sup>, esto no sucede. Aquí el interés por los demás es puramente instrumental o, al menos, no es necesario que suceda de otra manera.

Gauthier afirma que, al encontrar a los demás dispuestos a la cooperación, se desarrolla un interés por ellos y por la cooperación mutua. Pero ¿por qué? Los demás, al igual que yo, están dispuestos a cooperar porque en determinadas circunstancias eso les resulta ventajoso. ¿Por qué esto ha de hacerme tomar interés en sus asuntos? Si un colaborador esencial e insustituible cae enfermo lo sentiré mucho, pero sólo en tanto que eso afecta a la consecución de los objetivos comunes. A lo mejor me cae bien. Incluso es posible que, gracias a la participación en una tarea común, haya llegado a conocerle mejor y me sea simpático. Sin duda, a los seres humanos a veces les cae bien otro ser humano. Pero a veces no. Mi compañero, incluso cuando es insustituible, puede ser un pelmazo, o tener la fea costumbre de hablar con la boca llena y a gritos. Esto es igualmente cierto en los casos de las tare-

---

para conseguir en el regateo el mejor resultado posible. Pero esto no nos interesa ahora, porque la tarea colectiva, de la cual nos estamos ocupando, es bailar un paso a dos. La composición del cartel no es una tarea colectiva.

<sup>36</sup> En el caso de que algunos jugadores quieran la cooperación como fin y otros como medio esto no sucede. Pero tampoco en esos casos se plantearía el dilema (al menos en los juegos bipersonales y en los juegos n-personales en los que un número suficiente valore la cooperación como fin) puesto que la elección de A por parte de los jugadores que valoran la cooperación como fin sería suficiente para asegurar un resultado óptimo.

<sup>37</sup> Dijo que en la mayoría y no en la totalidad porque puede suceder que en una actividad colectiva lo único que uno busque principalmente sea la, digamos, parte colectiva, siendo más o menos indiferente a lo demás. Estos casos son semejantes a aquellos en los que la cooperación es el fin. Por ejemplo, yo puedo querer bailar con otra persona porque quiero bailar un paso a dos, pero también puede que bailarín paso a dos sea una excusa para bailar con otra persona.

as esencialmente participativas. Aunque querer bailar un paso a dos signifique querer bailar con otra persona, eso no me obliga a sentir ninguna simpatía por mi compañero de baile ni a tomarme ningún interés por sus asuntos más allá de que no se rompa un tobillo o se maree antes de la sesión de noche. Y este interés es el que siente un piloto por su coche. De modo que, aun en el supuesto caso de que fuera cierto que las actividades colectivas enriquecen la vida humana de un modo que las hace indispensables, de aquí no se sigue necesariamente que se genere espontáneamente ningún sentimiento altruista. Y respecto a que pase de encontrar la cooperación valiosa como medio a encontrarla valiosa en sí, depende fundamentalmente de lo anterior y no al revés como parece pretender Gauthier. No es que primero pase a valorar la cooperación en sí y después a los que cooperan conmigo. Más bien al contrario, si sucede que la cooperación me da ocasión de conocer a una persona que resulta que me es extraordinariamente simpática, y llego a tener interés no instrumental por sus asuntos, pasaré a encontrar en la cooperación un valor no dependiente sólo por el hecho que de me dará ocasión de seguir estando con esa persona unidos por una tarea común. Pero esto no sucede siempre. Hay gente con la que es un placer en sí participar en tareas colectivas y en todo tipo de actividades cooperativas. Con otros es un tormento.

Gauthier dice que estas afirmaciones son de carácter psicológico. Desde luego, lo son. Sólo que, por las razones expuestas, yo pienso que son dudosas. Más bien tiendo a pensar que en muchos casos es difícil que ocurra de ese modo, especialmente en las situaciones de dilema *n*-personal donde las relaciones de reciprocidad quedan perdidas en el anonimato. Pero en todo caso esta discusión queda fuera del alcance de este trabajo. Entonces, ¿qué es lo que en estas afirmaciones es realmente importante? Lo que se pretende al hablar del surgimiento de los sentimientos altruistas es una “reconstrucción racional”. Es decir, el propósito de Gauthier no es mostrar cómo surgen los sentimientos morales, sino mostrar que los sentimientos morales que de hecho tenemos (sea cual sea la forma en la que han surgido) tienen una justificación racional. Puesto que Gauthier parte de la TER, esto quiere decir que nuestros sentimientos morales se justifican racionalmente en tanto que su mantenimiento es un medio para la maximización de la utilidad o, al menos, si no es contradictorio con ella. Por tanto, el propósito de toda la argumentación es derivar la moralidad del interés propio. Ahora bien, ¿muestran esto los argumentos que hemos expuesto y analizado aquí? Lo que se ha pretendido mostrar es que

...una persona, reflexionando sobre sus sentimientos morales, los consideraría una extensión apropiada de su interés por los otros en el contexto de actividades participativas valiosas, y entonces consideraría esas actividades apropiadamente valiosas en tanto que están sujetas a constreñimientos morales.<sup>38</sup>

---

<sup>38</sup> Gauthier op. cit., p.339.

Nosotros hemos defendido que la cooperación no tiene por qué generar ningún sentimiento de interés por los asuntos de los demás. Esta afirmación no es psicológica. No dice nada acerca de cómo surgen o dejan de surgir los sentimientos morales. Lo que dice es que no hay nada en las tareas cooperativas, ni siquiera en la necesaria cooperación requerida por las actividades colectivas, que haga suponer que un agente racional encontrará que sus sentimientos morales (si los tiene) tienen en ellas su justificación, ni que haga pensar a un agente racional que carezca de tales sentimientos que haría bien en tenerlos.

La última parte del texto citado hace referencia a por qué deben ser valorados los constreñimientos morales. Si se pretende fundamentar la moral en el interés, la razón más directa sería que las constricciones morales, que el obrar moralmente, promociona el interés propio. Lamentablemente, esto no es así, salvo en el caso trivial de que estemos interesados en obrar moralmente. Otra posible razón es que los constreñimientos morales deben ser valorados porque hacen posible la cooperación y esta, a su vez, debe ser valorada en sí misma. Pero la validez de esta respuesta depende del supuesto de que la cooperación es valiosa en sí misma, supuesto que hemos rechazado. El único sentido en que esto puede mantenerse es en el de que un agente racional encontrará más valiosa la conducta cooperativa general que la no cooperación generalizada. Por ello, un agente racional estará dispuesto a cooperar sólo si su cooperación es condición necesaria para garantizar la cooperación de los demás. Pero como ya mostramos anteriormente, esto ocurre en contadas ocasiones.

Su argumento es que un agente racional, al encontrar valiosa la cooperación como medio, pasará también a encontrarla valiosa en sí. No sólo no hay ningún motivo para sostener esto, sino que, además, los dilemas surgen precisamente en los casos en los que ni siquiera se da a la cooperación valor instrumental. Salvo en muy determinadas circunstancias de reciprocidad directa e interacción repetida, un agente racional no encontrará valiosa la cooperación como medio. Y en aquellos casos en los que se valora la cooperación como medio, no parece haber motivos para suponer que se pase a valorar como fin. De modo que, aun aceptando que los constreñimientos morales hacen posible la cooperación, aquellos sólo se valorarán cuando se valore esta, lo que no sucede de modo necesario ni en la mayoría de los casos y, lo que es más importante, no sucede en los casos que presentan un problema, a saber, en las situaciones DP.

## **5. Restricciones morales**

Queda por saber, si, en el caso de que Gauthier hubiera demostrado lo que pretende demostrar, esto habría sido una derivación de la moralidad. Si el argumento fuera válido, se hubiera mostrado que un agente racional acaba por valorar la coo-

peración en sí y los intereses de sus compañeros junto con los propios. A pesar de que ese interés por el prójimo tendría su origen en el interés propio, lo trascendería y se haría autónomo, e.d., por ejemplo, haría que se tuviera interés por los demás incluso cuando la interacción cooperativa hubiera terminado. Es problemático que este interés se extendiera también a personas con las que nunca se coopera de hecho o a los que es imposible incluir en tareas cooperativas. Los habitantes de sociedades lejanas formarían parte del primer grupo. Los que no tienen nada que ofrecer formarían parte del segundo<sup>39</sup>. Para evitar estos problemas, supondremos una sociedad en la que todos entablan relaciones cooperativas con todos. En esta sociedad, cada miembro tiene un interés en el bienestar de todos los demás. En ella, las consideraciones de carácter moral comprometen no sólo la razón de los agentes, sino también (sobre todo) sus sentimientos. Esto es una forma de decir que los agentes elegirán siempre la estrategia A puesto que, debido al interés que sienten unos por otros, A no es peor para ellos. Gauthier afirma que una situación de este tipo puede tener dos orígenes distintos y que, por tanto, hay dos modos en los que la moralidad puede comprometernos afectivamente<sup>40</sup>.

En primer lugar, puede suceder que estas personas tengan capacidad para lo que podríamos llamar una *moral afectiva*, que no es sino una forma de altruismo universal. Es una moral basada en sentimientos de benevolencia. Pero si concebimos la moral básicamente como un conjunto de restricciones en la búsqueda del interés propio, salvo que definamos este de una forma meramente egoísta, la situación queda mejor descrita diciendo que, entre personas con este grado de benevolencia, la moral es innecesaria. Los intereses de cada uno están de tal modo determinados por consideraciones acerca del interés de los demás que, sin necesidad de constreñimientos de ningún tipo, todos elegirían siempre la estrategia A. Es decir, todos harían A porque sería la estrategia que maximizaría su utilidad. Que los intereses de cada uno sean altruistas no es más que un dato sobre el contenido de las preferencias, como lo es el que sean egoístas. La única diferencia es que entre este tipo de personas, siempre y cuando su altruismo sea moderado<sup>41</sup>, no se plantearían dilemas. Y no se plantearían porque no sucede que cada uno tenga un interés distinto del interés de los demás y que pueda entrar en conflicto con ellos. Más bien, todos tienen el mismo interés, a saber, que los intereses de todos se satisfagan de la forma más completa posible. Esto haría la moral innecesaria. Muchos autores han señalado que el supuesto de la benevolencia universal anula las condiciones que hacen necesario

---

<sup>39</sup> Una de las objeciones más comunes a la teoría de Gauthier es que deja fuera de la relación moral a todos aquellos con los que no es posible entrar en una relación cooperativa fructífera, tales como determinados inválidos físicos y psíquicos. Un ejemplo de estas críticas puede encontrarse en Calsamiglia "Un egoísta colectivo"

<sup>40</sup> Gauthier op. cit., p.327.

<sup>41</sup> Dos altruistas puros también se enfrentarían a dilemas, pues cada uno intentaría alcanzar un objetivo distinto y, de nuevo, contrapuesto: la utilidad del otro.

y posible el surgimiento de la moral. Por ejemplo, Hume afirma que la moral es necesaria porque a) los recursos con los que se cuenta para satisfacer los intereses son escasos y b) no existe la benevolencia universal<sup>42</sup>. Hay un sentido en el que la conducta basada en la benevolencia universal es en sí misma una conducta moral a pesar de no ser una conducta constreñida, pues nuestra conducta sería idéntica a la conducta que obedece a reglas morales. Concretamente, la existencia de esta moral afectiva haría que todos los agentes eligieran siempre la estrategia A. Pero esta coincidencia conductual no es suficiente para que podamos hablar de moral. Más bien lo que caracteriza la conducta moral, distinguiéndola de otros tipos de conducta, como admite Gauthier, al caracterizar la mora como un tipo de restricción en la maximización directa de la utilidad, es que la conducta moral representa una restricción. Cuando los agentes hacen suyos los intereses de los demás no se enfrentan a ningún conflicto de elección, no hay ningún conflicto entre A y otra conducta posible a la que llamamos S. No sucede que esta conducta altruista *se fundamente* en el interés propio. Es verdad que A es la estrategia justificada por el interés, pero sólo en el sentido trivial de que la conducta A *es al mismo tiempo* S.

Se podría objetar que lo que caracteriza la conducta moral es que los agentes tengan este sentimiento de benevolencia. Por ejemplo, Parfit sostiene algo parecido al pedir que el interés propio se defina de un modo independiente de las preferencias de los agentes. El propio Gauthier, que solía mantener una definición de auto interés muy cercano al de la Teoría de la Elección Racional estandar identificándolo con “cualquier valor” que quiera promocionar un agente, “quizá su propia felicidad o no”<sup>43</sup>, en los últimos tiempos parece haber modificado su postura, afirmando que “si me preocupo por ti, y por tanto encuentro bueno tu bienestar, no por ello se convierte en mi propio bien”<sup>44</sup>. Es cierto que no se puede hacer que el interés propio sea equivalente por definición a aquello que los agentes promocionan mediante su conducta. En este sentido, debemos rechazar afirmaciones del tipo “lo que cada uno hace es, por definición, lo mejor para él”<sup>45</sup>. Pero eso no significa que el interés propio deba definirse de una forma estrecha, entendiendo que “interés propio” hace referencia únicamente a intereses egoístas. Cuando una madre hace lo que es mejor para su hijo esta, por lo general, obrando interesadamente. Y si suponemos que el interés de los agentes benevolentes por los demás es un interés de este tipo, entonces cuando eligen A están obrando de un modo auto-interesado. Desde este punto de vista, la así llamada “moral afectiva” no se distingue de la conducta inte-

---

<sup>42</sup> Sobre las circunstancias que hacen posible el surgimiento de la moral ver Hume *Tratado de la naturaleza humana*, pp.708 y ss.

<sup>43</sup> Ver por ejemplo Gauthier “El egoísta incumplido”, p. 73.

<sup>44</sup> La cita está extraída de una conferencia pronunciada por Gauthier en Madrid el 19 de Mayo de 1998, con el título “The best of times”.

<sup>45</sup> Parfit *Reasons and Persons*, p.87.

resada. Esto es precisamente lo que se quiere decir al afirmar que la existencia de la benevolencia universal no sólo no es una forma de moral, sino que es una de las condiciones que hacen la moral innecesaria. Por otro lado, no resulta muy satisfactorio identificar la moral con algo que no está sujeto a la voluntad. Habitualmente entendemos que la moral se expresa en prescripciones que podemos obedecer. Y nadie puede obedecer una prescripción del tipo “ten sentimientos benevolentes”. Dejando aparte que la existencia de la benevolencia universal haga innecesaria la moralidad, está el hecho fundamental de que tal sentimiento no existe, salvo quizá muy excepcionalmente.

El segundo modo en que las consideraciones morales pueden comprometernos afectivamente es mediante una *capacidad afectiva para la moral racional*. Tener esta capacidad significa poder sentirse emocionalmente comprometido con aquello que reconocemos como una consideración moral válida. Esta capacidad, contrariamente a lo que sucedía en el caso anterior, presupone una concepción previa de la moralidad puesto que “uno no puede ser movido por un sentido del deber a menos que uno, con anterioridad, crea que alguna acción es su deber”<sup>46</sup>. Los agentes así dispuestos harían A independientemente de los beneficios que les proporcione, y sólo es una consecuencia posterior el que, al estar así dispuestos, A ya no resulte peor para ellos. Para Gauthier, esta actitud es la que propiamente merece el nombre de “moral”

Entendido en sentido propio, un hombre justo es aquel que, reconociendo un determinado curso de acción como justo, encuentra sus sentimientos comprometidos por ese reconocimiento y así se encuentra a sí mismo dispuesto a adherirse a tal curso de acción a causa de su justicia.<sup>47</sup>

No es mi intención contradecir a Gauthier en este sentido. A pesar de que a mi juicio es preciso hacer determinadas matizaciones a esta afirmación, creo que es básicamente correcta. La cuestión es si la moral así caracterizada puede basarse en el interés propio. Para poder contestar de modo afirmativo, sería necesario que el agente racional a) encontrara que la restricción moral de la conducta puede ser justificada en términos de interés propio, pues esto es la condición que hace que las restricciones morales sean racionales, y b) que posteriormente se encontrara afectivamente comprometido con tales restricciones. A lo largo de este trabajo hemos negado la primera de estas condiciones, asumiendo que, en las situaciones DP la conducta dictada por la razón es S y no A: la misma razón que nos aconseja llegar a acuerdos satisfactorios nos aconseja, una vez alcanzados que, en ausencia de mecanismos externos que hagan obligatorio su cumplimiento, desertemos siempre

---

<sup>46</sup> Gauthier op. cit., p.328.

<sup>47</sup> Gauthier *Ibid.*

y cuando tengamos unas expectativas razonables de no ser descubiertos y perder a largo plazo más de lo que ganamos con nuestra deserción. El motivo es que comparamos los beneficios obtenidos si nos atenemos al acuerdo con los del incumplimiento unilateral por nuestra parte. Sin embargo, Gauthier afirma en un texto posterior a *La moral por acuerdo* que la comparación que debemos hacer es otra<sup>48</sup>: un maximizador restringido comparará los beneficios obtenidos del cumplimiento mutuo del acuerdo con los que hubiera obtenido si el acuerdo no se hubiera establecido. Dicho de otro modo, un maximizador restringido tendrá como punto de comparación no la deserción individual sino el Estado de Naturaleza. Pero si el agente no tiene motivos para pensar que exista un peligro real de volver a la situación previa al acuerdo, ¿por qué tiene un motivo racional para atenerse al mismo? Volvamos a nuestro ejemplo ¿debo pagar al taxista, a pesar de que me beneficia más no hacerlo, sólo porque si no hubiéramos hecho un trato mi situación sería aún peor, incluso suponiendo que no vamos volver a encontrarnos, ni él tiene ningún medio de localizarme o identificarme ni de aplicar ninguna sanción? Gauthier reconoce que esta conducta sería, desde la Teoría de la Elección Racional ortodoxa, absolutamente irracional, y con ello reconoce que se ha deslizado a “una teoría alternativa que abarca la maximización restringida”<sup>49</sup>. Pero si bien no hay duda de que la teoría ortodoxa no contiene suposiciones morales entre sus premisas, no está tan claro que pueda decirse lo mismo de la teoría alternativa, a pesar de que Gauthier lo niega con firmeza.

Supongamos ahora que la condición a) se cumple y preguntemos si en tal caso se cumpliría la condición b) y se crearía un compromiso afectivo del agente racional con la moralidad. Podemos fijarnos en los casos en los que sí ocurre que la conducta cooperativa resulta beneficiosa en términos de interés, por ejemplo los casos de dilemas iterativos mencionados más arriba, en los que la situación DP se plantea a los mismos agentes en un número indeterminado de veces. En estos casos, un agente racional elegirá la estrategia A porque, sin necesidad de hacer ningún supuesto sobre el carácter moral de este agente, esta conducta le resulta a medio o largo plazo beneficiosa. Según Gauthier, un individuo liberal pasaría a encontrarse afectivamente ligado a la moral. “Sentirse afectivamente ligado a la moral” significa sentirse inclinado a realizar la conducta dictada por la moral debido a que esta es precisamente la conducta dictada por la moral. Pero suponer esta capacidad afectiva para la moral es innecesario, puesto que si la conducta moral beneficia al agente este se sentirá inclinado a ella de todos modos. Sin embargo, esto no hace justicia a Gauthier. La elección de A se justifica racionalmente porque es la conducta más beneficiosa a largo plazo, pero en cada caso concreto de elección, A puede resultar perjudicial. A pesar de esto, un individuo liberal hará A porque, una vez reco-

---

<sup>48</sup> Gauthier “Uniting separate persons”.

<sup>49</sup> Gauthier art. cit. p.186.

nocida la justificación de la moral en términos de interés, se siente emocionalmente comprometido con esta conducta. A su vez, este compromiso hace que A tampoco sea perjudicial en el caso concreto. Pero esto no significa que la moral sea trivialmente idéntica al interés propio, puesto que primero exigimos que la conducta moral tenga una justificación basada en el interés.

Aun entendido en este sentido, es posible pensar que el compromiso afectivo sigue siendo inoperante, pues un agente racional elegirá A en esos casos de todos modos porque A es la mejor conducta a seguir. En todo caso, el compromiso afectivo puede servir para garantizar la elección de A en los casos en los que el agente, debido, por ejemplo, a lo prolongado de la interacción, puede “olvidar” que A es beneficiosa. Un agente puede verse tentado a utilizar S por los beneficios inmediatos que esta conducta le reporta a pesar de que esto le perjudique a más largo plazo. Por eso, un compromiso afectivo con la moralidad, que nos disponga a elegir A con independencia del resultado, resulta útil. Pero es dudoso que Gauthier piense en estos casos, ya que los agentes racionales ideales en los que basa su teoría no tienen “olvidos” ni “tentaciones” de este tipo. El compromiso afectivo no sólo resulta inoperante sino que presenta cierta inconsistencia con la propia teoría de Gauthier. Si un agente sólo encuentra justificada la conducta moral cuando esta promociona el interés, no se entiende bien qué quiere decir que, una vez que ha visto la moral así justificada, está dispuesto a adoptar esta conducta con independencia de los resultados que esto tenga para él en términos de interés. Según Gauthier, el individuo liberal “no busca aprovecharse ni mejorar su situación con respecto a los demás procurando violentar unos constreñimientos justificados sin apelar al mutuo concernimiento.”<sup>50</sup> Pero si los constreñimientos morales están justificados en términos de interés, (que no lo están) malamente pueden ser violentados en el propio provecho, ya que por hipótesis lo que está en favor del propio interés es mantenerlos.

En resumen, Gauthier intenta dar una justificación racional, e.d., auto interesada, a la moralidad. Esta justificación sólo puede hacerse de dos maneras. O bien se entiende que estamos interesados en la moral, cosa que Gauthier descarta, o bien mostramos que obrar moralmente promociona nuestro interés, definido este de modo independiente. Esto es lo que Gauthier intenta. Pero, con las escasas excepciones mencionadas, esto no sucede. Simplemente, no es cierto que la virtud siempre (ni siquiera casi siempre) obtenga recompensa, y el caso en el que la virtud es su propia recompensa queda descartado por irrelevante. Pero si fuera cierto, lo que se mostraría es que no hay dilemas, no que haya una solución moral a los dilemas. Moral e interés coincidirían. Y esto, de nuevo, deja sin papel a la moral.

---

<sup>50</sup> Gauthier op. cit., p.329.

## 6. Entre Hobbes y Rousseau

La primera vez que leí un texto de Gauthier, hace ya algunos años, a pesar de sus repetidas referencias a Hobbes, percibí con claridad lo que podemos llamar un “cierto hábito roussonian” en su tratamiento del individuo liberal. Yo no me explicaba, y sigo sin explicarme, cómo de un individuo hobbesiano podía salir un individuo como el que Gauthier describía al final de su obra principal. Posteriores lecturas de Gauthier me hicieron confirmar mi impresión y también poder formularla en términos algo más precisos. Sería vano el intento de buscar en la caracterización de la razón el punto en el que Gauthier y Hobbes se separan, pues ambas se ajustan a lo que ha dado en llamarse *razón maximizadora*: una capacidad de “sumar y restar”, una facultad que “cuando concebía cualquier cosa, era capaz de investigar sus consecuencias y los efectos que podría producir con ella”<sup>51</sup> y que, como afirmaba Hume, sólo es y puede ser esclava de las pasiones. Debemos mirar por tanto a las propias pasiones.

El individuo hobbesiano es el mismo en el Estado de Naturaleza y después del pacto. Sus deseos, sentimientos y pasiones, así como su capacidad racional, no cambian. El pacto, que altera esencialmente la situación en la que se encuentra, no opera ningún cambio en él. Y, en la medida en que sigue siendo el mismo, necesita asegurar mediante mecanismos externos el cumplimiento de los pactos. El maximizador restringido de Gauthier puede hacer su aparición de una de estas dos maneras: o bien su racionalidad se defiende desde una “teoría alternativa” de la racionalidad, o bien el pacto opera en él una transformación moral que no involucra su razón, que sigue siendo la misma, sino sus sentimiento y pasiones dotándole de la mencionada capacidad afectiva para la moral racional. El punto más débil de la teoría de Gauthier es su empeño en mantener la primera línea de defensa, y al mismo tiempo ofrecer argumentos que en realidad están apoyando la segunda. Entrar a formar parte de una sociedad en la que sus miembros han llegado a acuerdos satisfactorios permite no sólo que los agentes obtengan mejor sus objetivos previos, sino también hace posible que surjan objetivos nuevos. Empleando la terminología que utilizamos antes, los acuerdos satisfactorios facilitan formas de cooperación ligadas a su objetivo por una relación causal y formas de acciones participativas ligadas a su objetivo por una relación formal. Estos objetivos nuevos y necesariamente ligados a la participación aumentan el catálogo de objetivos que los agentes pueden dese-  
ar<sup>52</sup>. Esto supone ya una primera incursión en territorio rousseauiano: se ambiciona lo que se ve, y la cooperación genera objetivos que solo entonces pueden ser deseados. En el Estado de Naturaleza no hay orquestas.

<sup>51</sup> Hobbes op. cit., p. 47.

<sup>52</sup> La importancia de estas nuevas posibilidades abiertas con la cooperación se señala especialmente en los trabajos de Gauthier más recientes, especialmente en “Contractualismo político”.

Sin embargo, esto no basta. Para que bastara habría que postular que tales objetivos son superiores y que son tales que un individuo racional debería querer alcanzarlos. Esta es la vía explotada por Gauthier y creo que sin éxito.

Hay otra vía que el propio Gauthier intenta en ocasiones<sup>53</sup>. Si no hay una transformación moral, las “leyes del soberano” incluso cuando son internalizadas, “son cadenas”. Esto no sucedería si pudiésemos aceptar libremente las restricciones morales a nuestra conducta. Gauthier no sigue este razonamiento hasta sus últimas consecuencias: seríamos por tanto libres, con un tipo de libertad (libertad civil) que reemplaza la libertad originaria (libertad natural) y es superior a esta. Si lo hiciera, se encontraría de bruces con el universo de Rousseau, en el que lo que se gana mediante la empresa cooperativa es tanto que el hombre “debería bendecir sin cesar el feliz instante que le arranco para siempre de aquella [la naturaleza] y que, de un animal estúpido y limitado, hizo un ser inteligente y un hombre”<sup>54</sup>. Lo más valioso de la cooperación no es por tanto la posibilidad de conseguir mejor cosas previamente deseadas, ni siquiera la posibilidad de poder plantearse objetivos nuevos, sino la propia transformación operada en los individuos que, al permitir contemplar las restricciones como emanadas de la propia voluntad, nos libra del sometimiento a la voluntad ajena y da a nuestras acciones “la moralidad que les faltaba antes.” Somos libres (y con capacidad moral) porque obedecemos a la voluntad general.

Lamentablemente, esta “vía roussoniana”, a pesar de ser más prometedora, sólo puede tener un éxito parcial. Para Gauthier, esta ampliación de horizontes acaba con el entorno de escasez que caracteriza a las situaciones DP y posibilita que la cooperación deje de verse como un mal necesario. Esto, sin embargo, está lejos de suceder y no porque la sociedad no sea lo que debe ser y no todos los acuerdos sean satisfactorios, sino por que la escasez es connatural al ser humano. Aunque en la sociedad haya más cosas y más valiosas, nunca habrá para todos todo lo que quieren. Junto con el aumento de las posibilidades aumentan los deseos y aparecen nuevas pasiones antes inéditas, y la posibilidad de obtener más de la deserción que del cumplimiento de los acuerdos siempre existe. Yo puedo reconocer que los acuerdos son plenamente satisfactorios, que es mejor tenerlos que no tenerlos, que mis posibilidades vitales son más ricas gracias a la existencia de empresas cooperativas y de la propia sociedad, en tanto que ella misma es una empresa cooperativa, y que sólo en ella llego a ser un ser humano en sentido pleno. Pero puede ser un ser humano defraudador. La transformación moral de la que hablan Gauthier y Rousseau hacen de mí un ser moral, pero no necesariamente un ser moral bueno: en el Estado de Naturaleza no hay injusticia, pero tampoco justicia; una vez establecidos los pactos, hay justicia, pero también injusticia.

Sería injusto decir que el razonamiento de la “vía roussoniana” es inoperante y

---

<sup>53</sup> Gauthier (93a)

<sup>54</sup> Rousseau *El contrato social* p.19

carece de validez. Antes bien, habla a favor de la propuesta de Gauthier de considerar que la comparación relevante para decidir si hemos de cumplir un acuerdo no es con lo que tendríamos si los incumpliéramos, sino con lo que tendríamos si tal acuerdo no existiera. Habla a favor de hacer un hueco a las consideraciones morales en nuestras deliberaciones. Al menos si somos honestos. Al menos si nos repugnan disfrutar de beneficios que no contribuimos a general. Pero al hablar así no apelamos (únicamente) a la razón, sino a unos sentimientos morales. Pero quizá esto no sea suficiente. Tener una voluntad general no impide que tengamos también una voluntad particular, ni ser ciudadanos impide que seamos también individuos particulares. El insensato de Hobbes<sup>55</sup> que, aun reconociendo un curso de acción como justo, piensa que la injusticia es compatible con la razón, puede hacer buenas migas con el insensato de Rousseau. Este puede incluso admitir que no cumplir con su parte sería “una injusticia cuyo progreso causaría la ruina del cuerpo político”<sup>56</sup>, ruina que en modo alguno desea. Pero si no progresa porque los demás cumplen su parte (o al menos si el incumplimiento propio no influye en el de los demás), su voluntad particular le inclinará a desertar de sus obligaciones.

Por eso, aun en una sociedad perfecta, en la que tanto la estructura social como los acuerdos concretos resulten satisfactorios, y a pesar de la transformación moral, la conclusión roussoniana sigue siendo hobbesiana: “Para que el pacto social no sea, pues, una vana fórmula, encierra tácitamente este compromiso (...) que consiste en que quien se niegue a obedecer a la voluntad general, será obligado por todo el cuerpo”<sup>57</sup>. Pero aunque el camino nos haya devuelto a Hobbes y a la necesidad e inevitabilidad de los mecanismos externos para asegurar el cumplimiento de los acuerdos, no lo hemos hecho en vano. Ahora sabemos que, con tales mecanismos, “se nos obligará a ser libres”.

## Referencias bibliográficas

- Axelrod, R.: *La evolución de la cooperación*, Madrid, Alianza Universidad, 1986.  
 Calsamiglia, A.: “Un egoísta colectivo”, *Doxa* nº6, 1989, pp. 77-94.  
 Francés, P. (ed.) *Egoísmo, Moralidad y Sociedad Liberal*, Barcelona, Paidós, 1989.  
 Gauthier, D.: “El egoísta incompleto”, en Frances (ed), *Egoísmo, Moralidad y Sociedad Liberal*, pp. 67-117.  
 Gauthier, D.: *Morals by Agreement*, Oxford, Clarendon, 1986.  
 Gauthier, D.: “Between Hobbes and Rawls” en Gauthier & Sugden (eds.), *Rationality, Justice and the Social*, pp. 24-39.

---

<sup>55</sup> Hobbes op. cit., p. 132

<sup>56</sup> El insensato de Rousseau aparece en op. cit., Libro I, capítulo VII.

<sup>57</sup> Rousseau op. cit., p.18

- Gauthier, D.: “Uniting separate persons” en Gauthier & Sugden (eds.), *Rationality, Justice and the Social*, pp. 176-192.
- Gauthier, D.: “Asegurar y amenazar” en Frances (ed), *Egoísmo, Moralidad y Sociedad Liberal*, pp. 119-159.
- Gauthier, D.: “Contractualismo político” en Frances (ed), *Egoísmo, Moralidad y Sociedad Liberal*, pp. 161-185.
- Gauthier, D. & Sugden, R. (eds) *Rationality, Justice and the Social Contract*, Michigan, University Press, 1993.
- Hume, D.: *Tratado de la Naturaleza Humana*, Madrid, Editora Nacional, 1981.
- Hobbes, T.: *Leviatan*, Madrid, Alianza Editorial, 1989.
- Kant, I.: *Fundamentación de la metafísica de las costumbres*, Madrid, Espasa-Calpe, 1946.
- Parfit, D.: “Prudence, Morality and the Prisoner’s Dilemma”, *The Proceedings of the British Academy*, Vol. LXV, 1979, pp. 539-564.
- Parfit, D.: *Reasons and Persond*, Oxford University Press, 1986.
- Rousseau, J.J.: *El contrato Social*, Madrid, Tecnos, 1988.

Blanca Rodríguez López  
Facultad de Filosofía  
Edificio A, Despacho 219  
Universidad Complutense de Madrid  
bmerino@filos.ucm.es