

## D. Gauthier o Hobbes sin Leviatán

JOSÉ MONTOYA  
(Universidad de Valencia)

Un tema clásico de la moderna filosofía moral ha sido el intento de reconciliación entre el interés propio y la conducta correcta. En la filosofía moral clásica el tema no se planteaba, o se planteaba de manera completamente distinta. El florecimiento del hombre, la *eudaimonia*, constituía en realidad el cumplimiento de sus verdaderos intereses, y la educación correcta enseñaba al joven a conocer y a sentir que los deberes formaban parte del sentido de su existencia. (No entraremos, naturalmente, a discutir los supuestos y la legitimidad de semejante concepción).

En todo caso, resulta claro que sólo con el nacimiento del moderno individualismo radical (con la destrucción de la idea del hombre como *animal social*) se ha desarrollado la creencia en la oposición por principio entre «intereses individuales» y «exigencias morales». Nacida de este individualismo radical, el problema de la moderna filosofía social es análogo al problema cartesiano: del mismo modo que, en el terreno de la teoría, Descartes pretendía deducir toda verdad de la única certeza del «*cogitans sum*», en el terreno de la práctica la moderna filosofía social pretenderá deducir cualquier exigencia moral del «*volens sum*»; y del mismo modo que, en el sistema cartesiano, las verdades así deducidas quedaban coloreadas en el sentido del idealismo por el método de su obtención, las exigencias morales así fundamentadas permanecerán afectadas por la duda de irrealidad. Si por *exigencia moral* entendemos la imposición por nosotros mismos de límites «internos» a nuestras acciones y a nuestros deseos, ¿cómo explicar esa adición entorpecedora al acervo de límites «externos» que nos son impuestos por la naturaleza y por la sociedad?

No intentaremos, desde luego, pasar revista a la serie de respuestas que se han dado a tal pregunta. En las líneas siguientes pretendemos tan sólo presentar una respuesta nueva y radical: la ofrecida por D. Gauthier en su obra *Morals by agreement*<sup>1</sup>. Gauthier no se inspira en su respuesta en las ideas humanitarias de Hume ni en las trascendentales de Kant, sino en las mucho menos respetables de Hobbes; pretende construir una moral de inspiración hobbesiana, que no emplee otra noción de racionalidad que la económica y que no necesite echar mano de sentimientos positivos de benevolencia y de desinterés<sup>2</sup>. Se trata de una moralidad de la *Weltklugheit*, de la prudencia mundana, con tal de que demos a este término un sentido no excesivamente mezquino.

1. D. GAUTHIER, *Morals by agreement*, Oxford: Clarendon, 1986.

2. Gauthier es un excelente conocedor de Hobbes, al que ha dedicado varios estudios, p. ej.: *The Logic of Leviathan: The Moral and Political Theory of Thomas Hobbes*, Oxford: Clarendon, 1969; «Hobbes's Social Contract», en: *Noûs*, 22 (1988), pp. 71-82; «Thomas Hobbes: Moral Theorist», en: *The Journal of Philosophy*, 1979, pp. 547-559; «Taming Leviathan», en: *Philosophy and Public Affairs*, 16 (1987), pp. 280-298.

Es generalmente conocido que Hobbes, en *Leviatán*, establece elegantemente las reglas morales fundamentales (las «leyes de la naturaleza») como máximas de prudencia. No expresan, para cada individuo, el comportamiento óptimo (que sería el ejercicio de una libertad omnimoda, basado en un dominio completo sobre los demás), pero constituyen un *second best* para todos y cada uno, pues son generalmente preferibles al estado de anarquía que se seguiría si cada uno buscara el ejercicio sin trabas de su libertad. (Prescindimos desde luego de los matices, a menudo sutiles, del razonamiento hobbesiano).

El problema para Hobbes consiste en asegurar el cumplimiento general de aquellas reglas; pues aún peor que el estado de anarquía sería para el individuo el estado de explotación: esto es, el estado en que él limitara su libertad y cumpliera las reglas, mientras que los demás no lo hicieran. Si no hay cumplimiento general, es preferible el estado de anarquía y de lucha de todos contra todos.

Para resolver este problema del cumplimiento de lo pactado, Hobbes —como es muy conocido— no encontraba otro medio que recurrir a una fuerza exterior (*Leviatán*, «ese dios mortal al que debemos, bajo el Dios inmortal, nuestra paz y defensa») dotada de los poderes más amplios posibles. Sólo un poder inapelable y omnipresente, el del Estado, puede garantizar el cumplimiento de las reglas de justicia y evitar así la recaída en el estado de anarquía.

Este punto de la ética hobbesiana, la teoría del cumplimiento, no es desde luego recogido por Gauthier. Si aceptamos ciertos supuestos plausibles acerca del funcionamiento de la sociedad, podemos prescindir de *Leviatán*, o al menos concederle una excedencia provisional.

El primero de estos supuestos es que nuestra sociedad se acerca razonablemente a la idea de un mercado perfectamente competitivo. No es, desde luego, un mercado perfecto; si lo fuera, no habría necesidad de imponer y de imponerse limitaciones morales (por razones que ya veremos). Sin embargo, se acerca razonablemente a él, y en ese sentido necesita de la moralidad como complemento, pero sólo como complemento: «la moralidad surge del fracaso del mercado».

¿Por qué un mercado perfectamente competitivo eliminaría la necesidad de la moralidad? La respuesta es simple si admitimos los principios básicos de la teoría económica, en especial el principio de utilidad marginal: el funcionamiento del mercado asegura idealmente la maximización de la utilidad de todos los participantes. Productores y consumidores intercambian mientras la producción y el cambio aumentan su utilidad respectiva. La situación en el mercado ideal es exactamente la inversa de la del dilema del prisionero, y por ello no es preciso ni conveniente la cooperación, si por «cooperación» entendemos la adopción de estrategias conjuntas. Justamente por ello, el mercado perfecto es una «zona moralmente libre», en la que las restricciones de la moralidad, impuestas por la necesidad de cooperar y de adoptar estrategias conjuntas, no tendrían ningún sentido.

Naturalmente, la existencia de un mercado perfectamente competitivo exige ciertas condiciones, que son en rigor imposibles de cumplir (razón por la cual el mercado perfecto debe considerarse como un ideal regulativo más que como una realidad posible). La primera es la ausencia completa de fuerza y de fraude: nadie es obligado a intercambiar, nadie es engañado con respecto a las cualidades de los bienes que intercambia. Se trata de una condición estructural, y que por tanto no debe confiarse a la simple buena fe de los individuos, sino que deberá estar garantizada por alguna institución civil o política.

Una segunda condición es la propiedad privada de todos los productos y factores de producción. Sólo bajo esta condición se produce el ajuste de las utilidades de las partes, pues sólo así se elimina el principal foco de distorsiones en el mercado: la existencia de externalidades. Hablamos de «externalidades» cuando un acto de pro-

ducción, intercambio o consumo, afecta —positiva o negativamente— a la utilidad de una persona que no participa en él. La instalación de una antena parabólica de TV en un bloque de viviendas, por ejemplo, afecta *externamente* de manera positiva al gozón (lo que no ha querido cotizar para su instalación: se trata de una externalidad positiva. Por el contrario, la polución del aire por una fábrica de cemento afecta negativamente a quienes habitan en los alrededores: una externalidad negativa. Como hemos dicho, la existencia de externalidades distorsiona gravemente los efectos maximizadores del mercado: afectan al precio real del producto (en el caso de externalidades positivas, quienes pagan, pagan más de lo que debieran; en el caso de externalidades negativas, menos de lo que debieran) y en consecuencia al suministro del bien en cuestión (los productos afectados por externalidades positivas tienden a ser ofrecidos por debajo de la demanda real, mientras que aquellos afectados por negativas tienden a ser ofrecidos en exceso). La eficacia maximizadora de la utilidad del mercado queda así gravemente limitada: algunos reciben más de lo que dan y otros menos.

Volvamos a la hipótesis del mercado perfecto, esto es, aquél en el que no existe coacción ni fraude, ni hay externalidades. El mercado así definido, piensa Gauthier, no hace más que potenciar enormemente los beneficios de la libre actividad del individuo. Un individuo aislado, Robinsón en su isla, utiliza sus capacidades (entendiendo por tales tanto sus dotes físicas y mentales como los medios que la naturaleza le ofrece) para satisfacer sus preferencias ordenadas. No podemos decir que ni sus capacidades, como recién definidas, ni sus preferencias limitan su libertad, sino hay que afirmar más bien que son las condiciones que la hacen posible.

Para Gauthier no hay una diferencia esencial, bajo este aspecto, entre Robinsón y el participante en el mercado. Ciertamente aquí las capacidades no están definidas tan sólo por la dotación mental del agente y las posibilidades del medio físico, sino también por la acción de otras personas (especificada a su vez por *sus* capacidades y *sus* preferencias); pero, mientras el mercado se mantenga plenamente competitivo, ello no afecta substancialmente a la situación. Simplemente, el agente se encuentra aquí con otro tipo de oportunidades, que incluyen las disposiciones de otras personas.

Hemos visto, por tanto, que el mercado, como pura idea de intercambio libre, es justo en cuanto aumenta la utilidad de todos los participantes y conduce a un óptimo paretiano. Ello no quiere decir, desde luego, que conduzca a un óptimo social y que sea equitativo en su sentido más profundo. Ello dependerá de la justicia del punto de partida: «los resultados del mercado son equitativos si, y sólo si, proceden de condiciones iniciales equitativas».

Las condiciones de equidad en el punto de partida pueden resumirse, según Gauthier, en la célebre cláusula de Locke: la apropiación de un bien es legítima cuando no perjudica a otros, es decir, «cuando queda aún bastante para los demás y tan bueno como lo que se toma». Lo que se quiere excluir es que exista una coerción injustificada en el punto de partida. Esta coerción existiría cuando un participante se presentara provisto de una hacienda adquirida tan sólo a costa de empeorar la situación de los demás. La cláusula de Locke lo que hace es introducir una estructura rudimentaria de derechos anterior incluso al funcionamiento del mercado, en lo que Gauthier llama «situación de interacción natural». El sentido de esta estructura de derechos consiste en autorizar a cada uno exclusivamente el uso de sus propios poderes, sin aprovecharse del ejercicio de las capacidades ajenas. La situación inicial determinada por el uso de estos derechos (es decir, por el ejercicio no coercitivo de los poderes de cada individuo) es una situación equitativa.

Si el mercado fuera perfectamente competitivo y la situación inicial equitativa, no había necesidad de moralidad, si por esa palabra entendemos la búsqueda de una cooperación positiva entre las personas, más allá del intercambio del mercado: «*morality arises from market failure*». También los utilitaristas clásicos, desde luego, habían

supuesto que la moral complementaba las deficiencias sociales, y que en una sociedad perfectamente organizada (es decir, estructurada por relaciones de intercambio realmente libres) no serían necesarias actitudes *específicamente* morales. Todos los ciudadanos serían, independientemente de su intención, agentes morales en cuanto, al buscar su felicidad en ese tipo de sociedad, contribuirían a maximizar la felicidad general. Ahora bien, los utilitaristas clásicos, siguiendo a Hume, encontraban en la benevolencia la raíz de aquella actitud moral que podemos llamar supletoria. Sólo un interés positivo por la felicidad general podría contribuir a eliminar las deficiencias de la organización social.

La postura de Gauthier en este punto es totalmente diferente. No cree que ni siquiera en este punto sea preciso eliminar esa condición necesaria de la racionalidad económica que es el no-tuismo, el desinterés recíproco de los participantes. Justamente aquí va a intervenir la idea de contrato, uno de cuyos elementos indispensables es precisamente el no-tuismo.

La respuesta racional al fracaso del mercado es la cooperación, y la cooperación es el dominio por excelencia de la justicia. Entraña la disposición a no sacar ventajas sobre otros que quizá el deficiente funcionamiento del mercado nos permitiría, sino más bien a actuar de mutuo acuerdo y a repartirse equitativamente la «plusvalía de utilidad» que la cooperación engendra frente al deficiente funcionamiento del mercado.

La cooperación no es una actitud positivamente benevolente, sino también racionalmente egoísta. No se cambian las motivaciones del agente económico, sino las circunstancias sobre las que esas motivaciones actúan. En efecto, cuando el mercado falla, nos encontramos en la situación retratada en el dilema del prisionero: si cada uno de nosotros busca maximizar su bien por su propia cuenta, todos nos encontraremos en peor situación que si nos decidimos a cooperar. Se trata de ponernos de acuerdo en cuál es el resultado de la acción conjunta más aceptable para todos, y después de aplicar cada uno la estrategia adecuada para lograr ese resultado.

Ahora bien, ¿cómo determinar qué resultados son aceptables para todos? La solución de Gauthier en este punto es innovadora y extremadamente técnica. Substituye la idea abstracta de contrato (sobre todo en el sentido rawlsiano de contrato para establecer la estructura básica de la sociedad) por la de negociación. La negociación supone posiciones iniciales, aspiraciones máximas que, si los agentes son racionales, están fijadas por lo que ellos estiman que es su contribución a la plusvalía de utilidad que resulta de la cooperación. Naturalmente, ninguna posición inicial será normalmente aceptable para todos los negociadores. Habrá por tanto que ponerse de acuerdo en alguna posición intermedia. Ahora bien, ¿cuál de las infinitas posiciones intermedias resultará la más aceptable para todos, y en este sentido la más racional? Sin duda, aquella que requiera una *concesión relativa mínima* por parte del negociador que más ceda. La presentación intuitiva es la siguiente: cada posición intermedia implica por parte de los negociadores una concesión. La importancia de esa concesión se mide en términos relativos: es decir, por la fracción del intervalo entre las posiciones iniciales extremas que esa concesión supone. Pues bien, entre el conjunto de posiciones intermedias posibles, debe seleccionarse aquella que minimice la concesión relativa del negociador que más ceda. (Simplificamos necesariamente: la explicación técnica, por elemental que se pretendiere, sería excesivamente larga). A este principio lo llama Gauthier «*principio minimax* de concesión relativa». En los casos más simples, cuando los frutos de la cooperación pueden ser tratados como un único bien transferible cuya cantidad no varía con la distribución, el principio minimax entraña por supuesto la distribución rigurosamente igualitaria de aquel bien.

Con el principio minimax de concesión relativa, Gauthier piensa haber establecido un criterio razonable de moralidad, o quizá mejor de justicia, en la distribución de los beneficios de la cooperación. Pero en una teoría completa de la moralidad no has-

ta con establecer el criterio de la moralidad, sino que hay que presentar también una teoría de la motivación moral (lo que los clásicos llamaban una teoría de la sanción). A esta teoría dedica Gauthier el capítulo VI de su libro.

Ya vimos cómo Gauthier no aceptaba una teoría de la motivación moral basada en el sentimiento de benevolencia o de humanidad. La motivación para observar los compromisos cooperativos no es, en opinión de Gauthier, sustancialmente distinta de la actuante en el mercado: la búsqueda del propio beneficio de acuerdo con la racionalidad económica. No cambia la motivación, pero sí la situación y en consecuencia también la estrategia. En lugar de seguir una estrategia individual, se abraza una estrategia conjunta porque, en esas circunstancias, es la más ventajosa para todos y cada uno. Ahora bien, esa estrategia conjunta no funcionará si, dentro de ciertos límites, no estamos decididos a seguirla de manera incondicional, incluso cuando su seguimiento entraña desventajas ocasionales para nosotros. De este modo, piensa Gauthier, dadas ciertas condiciones plausibles, un hombre verdaderamente racional, un maximizador de su propia utilidad, no dudará en asumir las limitaciones impuestas por el principio *minimax* y en obrar de acuerdo con ellas, incluso en ausencia de vigilancia. Por razones de maximización de su utilidad, el hombre racional renuncia, en las cuestiones negociadas, a actuar de acuerdo con el principio de maximización inmediata de la utilidad y decide simplemente cumplir con el acuerdo.

El hombre racional es así el hombre prudente con visión amplia <sup>3</sup>. Pero, precisamente por ello, no actúa de esa manera «altruista» si no se cumplen ciertas condiciones mínimas, que aseguren la obtención de la parte correspondiente de la «plusvalía» que la cooperación produce. Como Hume, Gauthier piensa que se dan ciertas circunstancias de la justicia, sin las cuales la observancia escrupulosa de las reglas de justicia carece de sentido, o al menos es ciertamente no racional. La más importante de ellas es, sin duda, la seguridad de un cumplimiento general, o al menos ampliamente extendido, de los acuerdos cooperativos. El maximizador racional está dispuesto a cooperar, incluso con ocasionales desventajas propias, si y sólo si puede razonablemente esperar que la mayoría de los participantes cumplan también los acuerdos, de manera que se produzca una efectiva mejora global en la condición de todos y de cada uno.

Ahora bien, esta confianza razonable en la cooperación general exige que haya en la sociedad una cierta medida de «traslucidez» (entre la absoluta transparencia y la opacidad completa) respecto al cumplimiento efectivo de los acuerdos. Gauthier intenta incluso establecer un índice numérico de translucidez, por debajo del cual deja de ser racional que la persona racional cumpla ella misma los acuerdos. No seguiremos esa tentativa. En todo caso, es evidente para Gauthier que en una situación de completa opacidad (es decir, cuando de ningún modo podemos estar seguros de que los demás cumplirán su parte del acuerdo y, por tanto, tampoco de que la observancia redundará en beneficio de todos y de cada uno) sería irracional mantener una actitud de estricta observancia. Por encima de todo, debemos estar seguros de que queda descartada la peor de las posibilidades, a saber, la explotación de quienes cumplen lo pactado por parte de quienes no lo cumplen.

Esta concepción de la moral (pues los acuerdos no se refieren tan sólo a distribuciones concretas de bienes sino a las reglas que han de regir la cooperación) puede parecer enormemente restringida, y Gauthier piensa que, de hecho, en lo que común-

---

3. En un artículo, publicado originalmente en 1967, y reproducido en castellano posteriormente («La moral y la ventaja», en: J. Raz (ed.), *Razonamiento práctico*, México: F.C.E. 1986), presentaba Gauthier una oposición entre prudencia y moral, que seguramente pretende haber superado en *Morals...*

mente se llama «moral» hay muchos otros elementos además del provecho racional, destacado por su teoría. Para Gauthier, la moral en el sentido que él propone (la búsqueda prudente de la propia utilidad en situaciones de cooperación) es simplemente la base o el suelo sobre el que descansa el desarrollo de la persona completa. Esta no es en la realidad un individuo aislado, que se limita a buscar su provecho de una manera racional, sino alguien que se relaciona con otras personas, que participa en actividades conjuntas, que valora esas actividades en cuanto permiten el desarrollo de cualidades que sólo en la colaboración pueden potenciarse, y que despliega libremente sus afectos hacia sus compañeros en esas actividades. En el contexto de actividades compartidas, las personas despliegan valores *tuísticos*, valores que van más allá de la moral estrechamente entendida. En el sentido defendido por Gauthier, la moral no expresa la preocupación de unas personas por otras, pero suministra las condiciones que alimentan dicha preocupación: «a just society provides a framework for community but is not communal».

Esta breve exposición no puede hacer justicia a la riqueza y densidad de argumentación que el libro ofrece. Habría que tener en cuenta, por lo demás, que Gauthier tiene un cierto sentido histórico del posible valor de lo que ofrece: su objetivo no es tanto establecer una teoría ética *sub specie aeternitatis* cuanto poner de manifiesto un aspecto importante de nuestra ideología moderna<sup>4</sup>. Sería sin embargo hacer poco honor al esfuerzo realizado el no presentar al menos algunas de las preguntas que su provocadora exposición suscita. Expondré sucesivamente, aunque quizá de forma no completamente articulada, las tres principales dificultades que la lectura de la obra me ha planteado.

La primera tiene que ver con la exigencia de ausencia de fraude y de coacción en el mercado perfectamente competitivo. Es obvio que Gauthier puede definir estipulativamente como estime conveniente la idea de mercado perfecto. Pero parece innegable que, si se estipula que en el mercado perfecto los participantes no ejercen fraude o coacción unos sobre otros, se está suponiendo implícitamente que el mercado perfecto está sometido a restricciones morales. Naturalmente, puede decirse que la ausencia de fraude y de coacción en el mercado no dependerían necesariamente de las disposiciones internas, «morales», de los participantes, sino de un aparato institucional (el Estado, probablemente) cuya única función consistiría en evitar la fuerza y el fraude de los intercambios libres. Pero, aparte de que ello significaría volver a introducir a Leviatán en la escena, resulta difícil pensar que un aparato exterior pudiera ejercer tal control sin la colaboración efectiva de gran parte de los agentes. No parece descabellada la sospecha de que el mercado perfecto no podría funcionar sin una fuerte dosis de moralidad en sus participantes. Y dado que los fallos del mercado se deberán, previsiblemente, a la falta de esa moralidad, ¿no es en cierto modo circular decir que la moralidad surge de los fallos del mercado?

La segunda pregunta tiene que ver con el concepto de negociación. Si he entendido bien la argumentación de Gauthier, no sólo cualquier distribución concreta de un bien, sino las mismas reglas de la moralidad (¡excepto las que prohíben el fraude y la violencia en los intercambios?) serían el resultado de un proceso de negociación regido por la concesión relativa *minimax*. Ahora bien, tengo serias dudas acerca de que las teorías de la negociación actualmente desarrolladas puedan acarrearse semejante peso. Pues toda la argumentación referente a la concesión relativa *minimax* depende

4. Véase a este respecto su artículo, «The Social Contract as Ideology», *Philosophy and Public Affairs*, 6 (1977), pp. 130-164.

de la fijación de las posiciones iniciales, a partir de las cuales se mide la amplitud de las respectivas concesiones; y nada nos garantiza que la fijación de esas posiciones iniciales sea razonable o equitativa. Lo más probable es que dependa de circunstancias coyunturales, como la posición relativa de fuerzas, la mayor o menor necesidad de llegar a un acuerdo, la existencia de reservas de las que mantenerse en espera de que el adversario ceda... Y sin una razonabilidad, o equidad, en la fijación de las posiciones de partida, no creo que pueda hablarse de la concesión relativa mínima como un criterio de justicia. Si queremos eludir el riesgo de abuso de fuerza en la misma negociación, y llegar a resultados equitativos, parece necesario recurrir, de una forma u otra, al artificio del velo de la ignorancia. Pero confieso paladinamente que esta dificultad puede fácilmente deberse a una deficiente comprensión por parte mía de la teoría de la negociación.

La tercera dificultad gira en torno a la teoría del cumplimiento o de la motivación moral. Francamente, no veo cómo, desde un punto de vista estrictamente racional, como creo que pretende ser el de Gauthier, se puede deslegitimar en principio la «free ridership». Al menos en muchas ocasiones<sup>5</sup> el «free rider» no perjudica en absoluto el éxito de las estrategias conjuntas, puesto que el efecto de su defección es absolutamente despreciable y (si el «free rider» es verdaderamente prudente) su defección, al permanecer oculta, no altera tampoco la voluntad de otros de seguir fielmente aquellas estrategias. La interiorización de las reglas de cooperación puede, sin duda, ser un buen instrumento para asegurar su cumplimiento extendido; pero el «free rider» (el «sensible knave» de Hume) estaría por encima de estas consideraciones y sabría en cada momento elegir la estrategia que maximizaría su utilidad sin poner en peligro la obtención de los beneficios derivados del seguimiento extenso de las estrategias conjuntas, del cumplimiento general de las normas morales. La única manera de evitar esta conclusión, me parece, es recurrir a sentimientos como honor, vergüenza, etc. para los que no estoy seguro de que la teoría de la racionalidad tenga una función específica.

Para concluir, quisiera dejar constancia de que el carácter fundamental de estas críticas (por otra parte, seguramente desorientadas) no empaña mi admiración por un esfuerzo tenso, continuado, ascético, de pensar, y por unas ideas que incluso los defensores de los más acendrados trascendentalismos harían bien en meditar.

---

5. Véase p. ej. el caso presentado por D. Lyons, *Forms and Limits of Utilitarianism*, Oxford: Clarendon, 1978, p. 163 s.