EL DILEMA DE GAUTHIER

Introducción

La tarea que se propone Gauthier es muy difícil de realizar, pero afortunadamente es mucho más sencilla de describir. Él pretende buscar un fundamento para la moral y cree haber encontrado ese fundamento en la racionalidad. Desafortunadamente, su teoría concluye en un dilema: si la conducta es racional no siempre es moral, y si la conducta es moral no siempre es racional.

Para mostrar el dilema voy a proceder de acuerdo a los siguientes pasos: primero, voy a examinar la caracterización de la racionalidad que propone Gauthier; segundo, trataré de mostrar que ese tipo de racionalidad no conduce a la aceptación de restricciones morales; tercero, analizaré el recurso empleado por Gauthier para arribar a restricciones morales y sostendré que no es racional aceptar ese tipo de restricciones; finalmente, estudiaré una posible solución del dilema, la que creo que no resulta exitosa.

La idea de la racionalidad

Según Gauthier, todas las personas son estratégicamente racionales y obran con un conocimiento completo de la situación en lo que respecta a los fundamentos para su elección[1]. Son también personas que no están interesadas en los intereses de aquellos con los cuales se encuentran en una relación de intercambio (pág. 87). Comenzamos entonces con personas que son racionales, conocedoras y desinteresadas. Pero todavía falta averiguar en qué sentido ellas son consideradas racionales.

[1] David Gauthier, *Morals by Agreement*, Oxford, Clarendon Press, 1986; todas las referencias que no contengan otra especificación se refieren a esta obra.

La racionalidad práctica, en su sentido más general, se identifica con la maximización y la cantidad a ser maximizada se asocia con la preferencia. Como a su vez la medida de la preferencia es la utilidad, la racionalidad se identifica con la maximización de la utilidad (pág. 22). Una primera idea de la persona racional, entonces, es la que considera que los seres racionales son aquellos que buscan la maximización de su utilidad.

No obstante, Gauthier reconsidera el concepto de preferencia distinguiendo en el mismo dos dimensiones: una se vincula con el comportamiento y la otra con la actitud. Si alguien elige una manzana de un plato que contiene manzanas y peras está revelando con su comportamiento su preferencia por comer una manzana. Pero también se expresa una preferencia por comer una manzana frente a una pera en la actitud de decir que se prefieren las manzanas a las peras (págs. 27-28). Y para que las preferencias puedan ser consideradas no debe haber un conflicto entre la dimensión del comportamiento y la dimensión de la actitud; las preferencias deben mostrarse estables bajo la experiencia y la reflexión (págs. 32-33). Una idea más ponderada de la persona racional indica ahora que ella es la que busca la maximización de sus preferencias consideradas y estables.

¿Cuál es la mejor manera de buscar esa maximización? Dadas ciertas condiciones plausibles y deseables, un maximizador racional de la utilidad enfrentado con la alternativa de no aceptar restricciones respecto de sus elecciones en la interacción o con la de aceptar ciertas restricciones que Gauthier especifica, elegiría esta última opción. Elegiría, de acuerdo a las bases maximizadoras de la utilidad, no hacer más elecciones sobre esas bases (pág. 158).

Lo que el individuo racional busca es la estrategia que le permita maximizar su utilidad en situaciones de interacción. Hay quienes piensan que es racional cooperar sólo si la utilidad que uno espera obtener actuando de acuerdo a la estrategia cooperativa conjunta es por lo menos igual a la utilidad que uno espera obtener aplicando la mejor estrategia individual (pág. 166). Gauthier rechaza esta idea puesto que piensa que una persona que actúa en base a una estrategia conjunta racional maximiza su utilidad, sujetándose a las restricciones impuestas por las utilidades que se conceden a todas las demás personas (pág. 167). La tesis de Gauthier es que -en contextos estratégicos- la disposición a realizar elecciones restringidas, en lugar de elecciones directamente maximizadoras, maximiza la utilidad (pág. 183). Para maximizar las expectativas de utilidad, entonces, hay que adoptar la estrategia de la maximización restringida (pág. 172).

El ejemplo más conocido, por supuesto, es el del Dilema del Prisionero:[2] si los dos individuos enfrentados al Dilema en su versión tradicional optan por restringir la maximización de su utilidad individual y buscan en cambio la cooperación recíproca, ambos logran maximizar su utilidad como resultado final.

Gauthier está proponiendo aquí una especie de defensa del utilitarismo de reglas frente al utilitarismo de actos. En ciertos casos la búsqueda directa de la utilidad es autofrustrante; es mejor buscar la utilidad indirectamente aceptando las restricciones impuestas por ciertas reglas, las que -al ser seguidas- producen realmente la máxima utilidad. Esto ocurre, por ejemplo, porque los maximizadores restringidos están en mejor posición que los maximizadores directos para inducir a otros a realizar con ellos transacciones mutuamente beneficiosas[3]. Un individuo que espera tener que interactuar con otros tiene una razón basada en la maximización de su utilidad para elegir convertirse en un maximizador restringido. Pero esta preferencia por la cooperación no está motivada por una preocupación hacia los demás, sino por su propio autointerés[4].

¿Cuándo decimos que el individuo *X* es racional de acuerdo a la teoría de Gauthier? Ante todo, el propósito de *X* es el de maximizar su autointerés. Podría hacerlo adoptando una estrategia maximizadora irrestricta, pero en ciertos casos esa estrategia resultaría autofrustrante, en el sentido de que ella no conduciría al resultado óptimo (así ocurriría en los casos del Dilema del Prisionero, por ejemplo). Entonces *X* decide adoptar una estrategia maximizadora restringida siempre que esas restricciones le aseguren una utilidad mayor que la que le procura la estrategia maximizadora irrestricta. Como se demuestra que así ocurre (otra vez el ejemplo típico es el del Dilema del Prisionero), *X* elige convertirse en un maximizador restringido. Pero no porque sus intereses se hayan modificado, por supuesto. Él sigue persiguiendo una sola cosa: la maximización de su utilidad. Sólo que ahora ha descubierto que ciertas restricciones a su conducta contribuyen

---

[2] Teniendo en cuenta los límites de este trabajo no me detendré a explicar el dilema. Remito a *Paradoxes of Rationality and Cooperation*, Richmond Campbell y Lanning Sowden (eds.), Vancouver, The University of British Columbia Press, 1985, y a Robert Axelrod, *The Evolution of Cooperation*, New York, Basic Books, 1984.

[3] Kurt Baier, «Rationality, Value and Preference», *Social Philosophy and Policy*, vol. 5, Issue 2, pág. 19.

[4] Edward F. McClennen, «Constrained Maximization and Resolute Choice», *Social Philosophy and Policy*, vol. 5, Issue 2, págs. 102 y 115.

a maximizar su utilidad, y por eso -y sólo por eso- las adopta.

Esta teoría de la racionalidad es generalmente aceptada. Parfit ha mostrado que la teoría del autointerés nos predispone muchas veces a actuar de un modo que, de acuerdo a la propia teoría, es irracional[5]. La idea de Gauthier acerca del maximizador restringido corrige ese defecto y se aproxima -de alguna manera- a la versión crítica de la teoría del propósito presente, que es la tesis que Parfit favorece[6]. No es una afirmación polémica, entonces, el sostener que el individuo *X* es racional. El problema interesante radica en investigar si un individuo que actúa sólo por las motivaciones expuestas es, además de racional, un individuo moral.

La idea de la moral

El individuo *X*, como he dicho, es indudablemente racional. Pero recordemos que él acepta restringir sus acciones sólo en la medida en que esas restricciones garanticen la maximización de su utilidad. ¿Es entonces *X* un individuo moral? Gauthier sostiene que la teoría moral es esencialmente la teoría de optimizar las restricciones sobre la maximización de la utilidad (pág. 78). Esta idea parece indicar que para que un individuo actúe moralmente basta con que acepte sólo este tipo de restricciones, y rechace en cambio cualquier restricción que minimice su utilidad.

Esto sí que es discutible. Como dice Morris[7] el autointerés es una característica de gran parte del comportamiento humano, pero no de aquella parte que es más admirable. Aun si definimos a la moral como la maximización restringida de la utilidad (y reconozcamos que no es un modo habitual de definirla) hay que especificar qué tipo de restricciones cuentan. Creo que nadie diría que *sólo* cuentan para caracterizar a la moral las restricciones que maximizan la utilidad a largo plazo. Atender a estas restricciones es, sin duda, racional, pero nadie sostendría seriamente la pretensión de que toda transacción racional es equitativa[8].

---

[5] Derek Parfit, *Reasons and Persons*, Oxford, Clarendon Press, 1984, pág. 13.

[6] Parfit, cit., págs. 117-120 y 125.

[7] Christopher W. Morris, «The Relation Between Self-Interest and Justice in Contractarian Ethics», *Social Philosophy and Policy*, vol. 5, Issue 2, pág. 119.

[8] Cfr. Jody S. Kraus y Jules L. Coleman, «Morality and the Theory of Rational Choice», *Ethics*, vol. 97, núm. 4, págs. 715 y 720.

Lo curioso es que el mismo Gauthier trata de caracterizar a la moral de un modo más estrecho. Por eso dice que coincide con Kant en que las restricciones morales deben aplicarse cualesquiera sean las preferencias individuales (pág. 100)[9] y observa también que las restricciones que surgen de la moral por convenio obligan racionalmente, en forma independiente de las preferencias particulares (págs. 328-329), de los deseos, propósitos e intereses del individuo[10]. Es que no todas las restricciones son restricciones morales; para serlo deben ser equitativas e imparciales[11].

Lo que se espera de un individuo moral es algo más que un cálculo acerca de la maximización de su utilidad. En otras palabras, lo que se espera de un individuo moral no es lo mismo de lo que se espera de un individuo puramente racional. Un individuo moral debe actuar -o abstenerse de actuar- sobre la base de ciertas restricciones que, en algunos casos, no maximizan su utilidad. No es sólo que el individuo moral deba actuar -u omitir- absteniéndose de calcular su utilidad. Es que debe actuar -u omitir- incluso después de haber calculado su utilidad y de haber descubierto que esa restricción en particular la disminuye.

Rawls ha expuesto con toda claridad el caso de la moral. Tener una moral implica por lo menos el reconocimiento de principios que se aplican imparcialmente, tanto a nuestra conducta como a la de los demás, principios que pueden constituir una restricción, o una limitación, respecto de la persecución de nuestros propios intereses. Uno debe reconocer los principios de la moral incluso cuando nos son desventajosos. Un hombre cuyos juicios morales coinciden siempre con sus intereses puede ser sospechoso de no tener moral alguna[12].

De modo que hasta ahora lo que hemos logrado caracterizar es al individuo racional. Este actúa en base a un cálculo que tiende a la maximización de sus preferencias consideradas y estables, esto es, de su utilidad. Aceptando ciertas restricciones esa utilidad aumenta y por eso el individuo racional decide aceptarlas. Pero como esas restricciones se aceptan sólo porque ellas maximizan la utilidad, y no se aceptarían si la disminuyera, no

---

[9] David Gauthier, «¿Por qué contractualismo?», *Doxa*, núm. 6, pág. 23.

[10] Gauthier, «¿Por qué contractualismo?», cit., pág. 20.

[11] Kraus y Coleman, cit., pág. 722.

[12] John Rawls, «Justice as Fairness», *Philosophy, Politics and Society*, Second Series, Peter Laslett y W. G. Runciman (eds.), Oxford, Basil Blackwell, 1972, pág. 139.

puede decirse que ese individuo actúe moralmente. (Y puesto que las restricciones maximizan la utilidad incluso es interesante considerar en qué medida pueden denominarse «restricciones».) Estamos ante el primer cuerno del dilema que enuncié al comienzo: el individuo *X* es racional pero no es moral.

La introducción de restricciones morales

Claro que, a esta altura de la cuestión, podría formularse una objeción obvia: la caracterización anterior es interesante pero *no es* la que propone la teoría de Gauthier. Me apresuro a conceder que esta objeción es parcialmente exacta. Gauthier procura introducir ciertas condiciones que transformen la situación que he descrito en una situación moral. Reconoce expresamente que puede ser acusado de probar que es racional aceptar restricciones, pero que esas restricciones pueden plausiblemente no ser consideradas morales[13]. Voy a examinar los esfuerzos de Gauthier para remediar esta situación, los que -a mi criterio- concluyen conduciendo al segundo cuerno del dilema.

Lo que Gauthier desea mostrar no es sólo que ciertas restricciones maximizan la utilidad; él quiere mostrar que los únicos acuerdos que pueden concertarse racionalmente son los acuerdos equitativos. Por eso la posición inicial a partir de la cual los acuerdos se celebran es una posición corregida, una posición en la cual ciertas ventajas son removidas, una posición en la que se respeta la versión de Gauthier de la «condición lockeana»[14]. Claro que la versión que Gauthier proporciona de esta condición es excepcionalmente permisiva, pero aun así ella implica la introducción de ciertas restricciones que no tienden a la maximización de la utilidad. Él interpreta a la condición de Locke como prohibiendo empeorar la situación de los otros, excepto cuando esto sea necesario para evitar empeorar nuestra propia situación (pág. 203). En la interacción natural debemos preguntar qué puede tomarse como punto básico para determinar los efectos que tienen las acciones de una persona respecto de la situación de otra, o de su propia situación. La distinción crucial que debemos

[13] Gauthier, «¿Por qué contractualismo?», cit., pág. 21.

[14] También Gauthier establece otro tipo de restricción al proponer su principio distributivo, denominado «concesión relativa minimax». Yo creo que el principio no está exento de dificultades, pero su estudio excede los límites de este trabajo. Remito a mi comunicación al *Lateinamerika und Europa im Dialog*, Münster, 1987, titulada «Recent Critiques of the Difference Principle».

establecer es entre empeorar la situación de algunos y fracasar en mejorarla; la condición prohíbe lo primero pero no lo segundo. El punto básico para determinar cómo yo afecto a otro, está determinado por el resultado que el otro esperaría en mi ausencia. El empeorar y el mejorar son juzgados comparando lo que yo hago con lo que podría haber ocurrido -*ceteris paribus*- en mi ausencia (págs. 203-205).

Por lo que puede verse, Gauthier interpreta la condición lockeana de modo que ella prohíbe empeorar la situación de otra persona interactuando con ella, salvo para evitar empeorar nuestra propia situación. Prohíbe mejorar la situación de uno a través de una interacción que empeore la situación de otro y expresa la idea subyacente de que no se tomen ventajas.

El papel primario de la condición lockeana es hacer posible las estructuras posteriores que son requeridas para las formas de interacción social. Ella convierte las libertades ilimitadas de la naturaleza hobbesiana en derechos y deberes exclusivos. Pero la condición lockeana, en la versión de Gauthier, no dice nada acerca de la igualdad, ni acerca de la satisfacción de las necesidades (págs. 205, 208-209 y 218).

Acepto que la interpretación de Gauthier de la condición lockeana es tan restrictiva que difícilmente pueda ocasionar grandes cambios respecto de la distribución de la propiedad y del ingreso en el mundo real. Pero no puede discutirse que es susceptible de provocar *algunos* cambios y de introducir restricciones que no resulten maximizadoras de la utilidad en el caso de algunos individuos. Por eso Gauthier no habla de justificar acuerdos reales entre individuos sino de justificar acuerdos hipotéticos, acuerdos en los que se supone que se respeta su versión de la condición lockeana. Son esos acuerdos los que proporcionan un criterio de justificabilidad de las prácticas morales existentes[15]. Aunque en el mundo real comenzamos con un conjunto existente de prácticas morales entendidas como restricciones a nuestro comportamiento maximizador, somos conducidos mediante la reflexión a la idea de un conjunto modificado que obtendría el acuerdo de todos. La capacidad reflexiva de los agentes racionales los conduce de lo dado a lo convenido, de las prácticas existentes y de los principios que requieren restricciones a aquellos que obtendrían el consentimiento de todos los individuos[16].

---

[15] Gauthier, «¿Por qué contractualismo?», cit., pág. 29.
[16] Gauthier, «¿Por qué contractualismo?», cit., pág. 32.

Ahora sí nos encontramos dentro del ámbito de la moral. No se trata ya de aceptar sólo aquellas restricciones que maximicen nuestra utilidad. Se trata, en cambio, de aceptar todas las restricciones que sean necesarias para lograr el acuerdo de todos, aunque la utilidad de varios de los participantes del acuerdo resulte disminuida. Las restricciones no coinciden ya con la utilidad, toman en cuenta a la imparcialidad y a la equidad, y poseen las características que solemos atribuir a los principios morales, tal como los he descrito más arriba.

El problema, obviamente, consiste ahora en mostrar la racionalidad de aquellos individuos que deben resignar propiedades o ingresos como consecuencia de este nuevo tipo de restricciones. Al fin de cuentas, el individuo puramente racional debe actuar inmoralmente, por lo menos algunas veces[17]. Porque una cosa es establecer lo racional de la autorrestricción y otra es establecer que el cumplimiento de un acuerdo es racional sólo si los términos de ese acuerdo son equitativos[18]. El propio Gauthier advierte este problema. Por eso observa que sería irracional para cualquiera el abandonar los beneficios que le proporciona el orden moral existente simplemente porque comprende que este orden le proporciona más de lo que él podría esperar de un acuerdo puramente racional[19]. Buchanan es igualmente escéptico respecto de las modificaciones a la situación real: «El ejercicio contractualista -dice- no requiere la rectificación de injusticias pasadas para poder dedicarse a resolver las cuestiones futuras relevantes»[20].

Para resolver este problema Gauthier recurre a la idea de la estabilidad social: una sociedad que sólo acepta acuerdos equitativos es una sociedad estable. En caso contrario, aquellos que se encuentran en una posición privilegiada únicamente pueden mantenerla recurriendo a la coerción[21]. Pero, ¿por qué sería irracional recurrir a la coerción para resguardar un privilegio, cuando esto puede hacerse con buenas posibilidades de éxito? Supongamos que los negros fueran minoría en Sudáfrica y los blancos rehusaran renunciar a sus privilegios, prefiriendo mantenerlos en base a la coerción. Siendo los blancos mayoría existen buenas probabilidades de que la coerción sea efectiva. En ese

[17] Kraus y Coleman, cit., pág. 715.
[18] Kraus y Coleman, cit., pág. 740.
[19] Gauthier, «¿Por qué contractualismo?», cit., pág. 33.
[20] James M. Buchanan, «The Gauthier Enterprise», *Social Philosophy and Policy*, vol. 5, Issue 2, pág. 87.
[21] Gauthier, «¿Por qué contractualismo?», cit., pág. 34.

caso los blancos sudafricanos se estarían comportando de un modo inmoral, pero no de un modo irracional. Más aún: si la coerción fuera seguramente eficaz, los blancos se estarían comportando de un modo irracional si decidieran renunciar a sus privilegios por razones morales. Porque este proceder contrariaría sus preferencias consideradas y estables acerca de la maximización de su utilidad.

De modo que ahora hemos logrado caracterizar al individuo moral. Este actúa aceptando restricciones que no siempre maximizan sus preferencias consideradas y estables; acepta -en cambio- restricciones que son equitativas e imparciales. Estamos ahora ante el segundo cuerno del dilema enunciado más arriba: el individuo *X* es moral pero no es racional. Y ahora no puede argüirse que esta no es la propuesta de Gauthier porque ella se ajusta exactamente a su teoría.

La presunta solución del dilema

Hay un recurso a nuestra disposición para intentar evadirnos del dilema, pero no es un recurso satisfactorio. Gauthier pretende fundar la moral en la racionalidad, y para solucionar el dilema podemos intentar distinguir dos tipos distintos de racionalidad. Uno de ellos es el de la racionalidad como inteligibilidad. Consiste en una mínima coherencia que se le exige al individuo y funciona sólo con un carácter explicativo. El otro tipo es el de la racionalidad como justificación. En este caso se exigen razones adecuadas, pero el acto no sólo es explicado sino también justificado[22]. En el caso de las restricciones que maximizan la utilidad el individuo es racional en el sentido explicativo, en el sentido de la racionalidad como inteligibilidad. En el caso de las restricciones equitativas e imparciales que no maximizan la utilidad el individuo que las acepta es racional en el sentido de la racionalidad como justificación.

Pero esta presunta solución es inaceptable. Porque mientras el concepto de racionalidad como inteligibilidad es moralmente neutro, la racionalidad como justificación está cargada de nociones morales. Las «razones adecuadas» que este último concepto exige son adecuadas -precisamente- desde el punto de vista moral. Y no puede pretenderse justificar racionalmente a la

[22] Josheph Mendala, «Gauthier's *Morals by Agreement* and Two Kinds of Rationality», *Ethics*, vol. 97, núm. 4, pág. 769.

moral con una noción de racionalidad que contiene -ella misma- rasgos morales. Aunque esto pudiera hacerse sin circularidad (y no veo cómo podría lograrse), de todos modos, el intento contrariaría la propia idea de Gauthier de hallar un modo no-moral de justificación, una justificación que él llama «deliberativa» o «neutral»[23]. Contrariaría su intento de definir la racionalidad independientemente de la moral[24].

Conclusión

Gauthier se propuso fundamentar la moral en la racionalidad y fue rápida presa de un dilema. Si caracterizaba adecuadamente a la racionalidad, el individuo racional no aceptaba restricciones morales; y si caracterizaba adecuadamente a la moral, la conducta del individuo moral no siempre era justificable racionalmente. El dilema no se soluciona distinguiendo dos conceptos distintos de racionalidad, puesto que la racionalidad como justificación no es un concepto moralmente neutro. El dilema planteado al comienzo sigue en pie: si se es racional no siempre se es moral, y si se es moral no siempre se es racional. La tarea que Gauthier se propuso realizar es de dimensiones gigantescas y no debe extrañar que fracasara en su intento. En realidad, lo sorprendente es que haya aportado argumentos tan buenos como para llegar muy cerca del éxito en una empresa tan complicada.

[23] Gauthier, «¿Por qué contractualismo?», cit., pág. 24.
[24] David Gauthier, «The Unity of Reason», *Ethics*, vol. 96, núm. 1, pág. 85.