CRÍTICA, Revista Hispanoamericana de Filosofía. Vol. 37, No. 110 (agosto 2005): 79–104

LA MORAL Y SUS SOMBRAS: LA RACIONALIDAD INSTRUMENTAL Y LA EVOLUCIÓN DE LAS NORMAS DE EQUIDAD

ALEJANDRO ROSAS
Departamento de Filosofía
Universidad Nacional de Colombia
arosasl@unal.edu.co

RESUMEN: Los sociobiólogos han defendido una posición "calvinista" que se resume en la siguiente fórmula: si la selección natural explica las actitudes morales, no hay altruismo genuino en la moral; si la moral es altruista, entonces la selección natural no puede explicarla. En este ensayo desenmascaro los presupuestos erróneos de esta posición y defiendo que el altruismo como equidad no es incompatible con la selección natural. Rechazo una concepción hobbesiana de la moral, pero sugiero su empleo en la interpretación de la psicología de los primates no humanos y en un modelo de progresión evolutiva que habría llevado a la moralidad como adaptación pasando por la razón instrumental.

PALABRAS CLAVE: sociobiología, altruismo, egoísmo, Frans de Waal

SUMMARY: Sociobiologists have endorsed a "Calvinist" position captured in the following formula: if natural selection explains moral attitudes, morality is not genuinely altruistic; if morality is altruistic, then natural selection cannot explain it. I expose the false presuppositions behind this claim and argue that altruism as fairness is not incompatible with natural selection. I reject a Hobbesian view of morality as an instrumental endorsement of fairness norms, but suggest its use to interpret primate psychology and to model an evolutionary progression ending in moral capacities as adaptations.

KEY WORDS: sociobiology, altruism, selfishness, Frans de Waal

1. *Sociobiología calvinista*

En su conferencia "Evolution and Ethics" (1995/1893), Thomas Henry Huxley protestaba contra la indiferencia moral de la naturaleza y sus productos, y sostenía: "Entendamos de una vez por todas que el progreso ético de la sociedad consiste, no en imitar el proceso cósmico [de la evolución], mucho menos en huir de él, sino en combatirlo" (1995, p. 134).[1] Esta tesis sobre la oposición entre evolución y moralidad adquirió nueva vigencia a partir de que los biólogos comenzaron a mirar la evolución y la selección natural desde la perspectiva del gen. George C. Williams, reconocido por su libro *Adaptation and*

[1] Todas las traducciones son mías.

Natural Selection (1966) como uno de los padres de esta nueva perspectiva, considera que ella otorga un renovado soporte al juicio de Huxley: "Nuestro concepto moderno de la selección natural puede describirse honestamente como un proceso que maximiza el egoísmo miope" (Williams 1995/1988). La moralidad no puede ser sino "una capacidad accidental producida gracias a la ilimitada estupidez de un proceso biológico que normalmente se opone a la expresión de tal capacidad" (citado en de Waal 1996, p. 2). George Williams ha reunido para sus lectores un deprimente catálogo de observaciones etológicas sobre asesinato, violación, infanticidio y canibalismo en el mundo animal (Williams 1995, 1989).

En su opinión, la naturaleza no simplemente da sobradas muestras de indiferencia moral, sino más bien de descarada inmoralidad. Williams señala que la serpiente cascabel tiene armas "precisamente diseñadas y usadas para producir víctimas" (1995, p. 319) y que el dolor y la muerte provocados por el uso agresivo de esas armas naturales debe distinguirse del generado por un proceso natural ciego, como el golpe de un rayo. Aunque la mente de una serpiente no es comparable en imputabilidad a la de un humano, G.C. Williams asume que entre ambas hay mayor cercanía que entre la mente animal y la total carencia de intencionalidad de un proceso físico. Williams no plantea el espinoso tema de la intencionalidad animal, pero da por supuesto que los animales la poseen en grado suficiente como para que se justifique hablar de inmoralidad en este caso, pues también hablamos de la inmoralidad de la guerra, aunque sabemos que surge por causas complejas y que, en muchos casos, no podemos imputar a nadie en particular la plena responsabilidad por ella.

Este punto oscuro pero crucial de la intencionalidad siguió proyectando su sombra sobre la comprensión moral del mundo diseñado por la selección natural. Cuando Richard Dawkins describió provocadoramente la perspectiva del gen como unidad de selección en *El gen egoísta* (1989/1976), hizo explícito que "egoísmo" aplicado a los genes o a los comportamientos por ellos codificados es un término técnico que se define por los efectos sin ninguna referencia a motivos o intenciones; pero no se atuvo consistentemente a esta convención y se dejó tentar por usos psicológicos del término. Enfatizando que los genes son egoístas en el sentido de que sus efectos producen beneficios a *corto plazo* para copias de sí mismos (el egoísmo *miope* al que se refiere Williams en el pasaje antes citado), Dawkins alude a la capacidad humana para la prudencia y para el altruismo, es decir, para enfocarse en los beneficios futuros o ajenos, como capacidades con las que nos "rebelamos contra la tiranía de los replicantes egoístas" (Dawkins

1989, p. 201). La sugerencia escondida en estas palabras es que los genes nos empujan de suyo a un egoísmo psicológico miope, pues de lo contrario no habría razón para ver las capacidades psicológicas para la prudencia y el altruismo como una rebelión contra los genes.

Con esa frase, Dawkins retoma la tesis de la oposición entre ética y evolución; pero no esgrime ningún argumento que defienda la tesis según la cual el "egoísmo" de los genes, en un sentido evolutivo y técnico, conduce al egoísmo psicológico. Su postura oficial es negar cualquier compromiso con una tesis de esa naturaleza, pues insistió en que usaba términos como "egoísmo" y "altruismo" en sentido técnico, sin implicar intenciones: "No me ocupo aquí de la psicología de los motivos" (Dawkins 1989, p. 4). Pero estas declaraciones no le impiden conectar, en la referencia del párrafo anterior y sin justificación alguna, la perspectiva sociobiológica del gen egoísta con la idea de una oposición entre moral y naturaleza. "Sociobiología calvinista" es, por esta razón, un apodo apropiado para esta perspectiva, apodo propuesto por el primatólogo holandés Frans de Waal (1996, p. 13).

2. *Genes egoístas y fenotipos altruistas*

La teoría del "gen egoísta" afirma que la evolución sucede en gran medida gracias a un proceso de selección natural que opera sobre los genes. Sólo los genes que se benefician a sí mismos o a sus copias sobreviven a este proceso. Se trata de una teoría verosímil, aunque controvertida; no es mi propósito aquí cuestionarla. Supongamos que es correcta y que un fenotipo adaptativo es aquel que ayuda al alelo que contribuye a su producción a reproducirse con más éxito que los alelos alternativos. El alelo es "egoísta" porque obtiene una ventaja reproductiva a través de los fenotipos que contribuye a producir. Obviamente, el alelo no lo hace de modo intencional. Es simplemente un hecho bruto, al que llamamos selección natural, que los alelos o bien son egoístas en el sentido explicado, o bien desaparecen. Pero, ¿hay razón para sostener que el fenotipo producido por genes egoístas también debe ser "egoísta" de manera tal que si se trata de un fenotipo psicológico, será egoísta en el sentido psicológico del término?[2]

[2] En su sentido psicológico, el término 'egoísmo' se usa ambiguamente en la literatura científica y filosófica. Sober y Wilson (1998) se han esforzado por introducir un uso preciso para discusiones que relacionen evolución y psicología. En lo que sigue entenderemos que 'egoísmo' se aplica, en primer lugar, a los motivos de una acción, y en particular a aquellos cuyo contenido proposicional menciona de manera irreductible un beneficio para el sujeto del motivo así calificado; mientras que, si

El desliz de Dawkins traiciona una inferencia directa de este estilo; pero es una inferencia inválida, debido a la diferencia existente entre las causas últimas y las causas inmediatas o próximas de un comportamiento (Sober 1993, pp. 197–198). La causa última de la evolución y existencia de un rasgo *C* es el efecto diferencial del rasgo *C* en el éxito reproductivo de la entidad, cualquiera que sea, que garantiza causalmente la existencia de *C* en los descendientes. Si *C* es un comportamiento, podemos asumir que beneficia, en primer lugar, al gen (tipo, no particular) que es la causa detonante (no la única) de la presencia de *C*. Por lo general, el comportamiento beneficia también al individuo que porta el gen; pero, en ocasiones, el comportamiento puede beneficiar a copias del gen que se encuentran en otros individuos. Son esos casos los que impiden que podamos pasar simplemente del "gen egoísta" al comportamiento egoísta. Este paso es especialmente problemático si, en lugar de centrarnos en el comportamiento, nos enfocamos en las causas próximas que los producen, atendiendo a la distinción de Sober ya mencionada.

El caso crucial es el de los organismos con mentes complejas, cuyos comportamientos se explican por causas próximas del tipo de las actitudes proposicionales (pienso que. . . , creo que. . . , deseo que. . . , quiero que. . .). Hablamos de motivos justamente cuando se trata de organismos con actitudes proposicionales como creencias y deseos. En estos casos, un comportamiento evoluciona cuando beneficia al gen que lo codifica y a sus copias; pero ello no implica que los *motivos* del comportamiento sean egoístas, es decir, que estén dirigidos a un beneficio personal como propósito último e irreductible de la acción. Hay al menos dos ejemplos en los que más bien lo contrario es posible. Se trata de casos en los que una motivación altruista genera, mejor o más confiablemente que una egoísta, los comportamientos necesarios para que el gen, e incluso el individuo que lo porta, promuevan su propia reproducción. En estos casos, el gen debe codificar comportamientos altruistas para ser un auténtico gen "egoísta".

El primer ejemplo es el del cuidado parental. A menudo, el cuidado de los críos exige que un progenitor tenga que posponer su propio bienestar. Si el progenitor no cumple esa exigencia, seguramente tendrá menos descendientes que otro que sea capaz de cumplirla. Especialmente en las especies en las que los críos son inicialmente

menciona un beneficio para un sujeto distinto del sujeto del motivo, este beneficio se toma como medio para un beneficio egoísta. Altruistas son los motivos que no son egoístas en este sentido. Véase también Rosas 2002.

incapaces de sobrevivir por sí solos, como en la humana, es natural suponer que el altruismo psicológico hacia los hijos es requerido por los "genes egoístas". En estas especies, los genes que programen un comportamiento *psicológicamente* egoísta respecto de la prole serán desplazados o sustituidos por genes que programen un comportamiento altruista, pues el cuidado parental de los altruistas es de mayor calidad y mejora su éxito reproductivo (Sober y Wilson 1998). Objetar que los progenitores son egoístas cuando atienden al bienestar de sus hijos porque garantizan así la reproducción de sus propios genes es pasar por alto la diferencia entre el sentido técnico de "egoísmo" en la biología evolucionista y el sentido psicológico usual.

El segundo ejemplo es más controvertido, pero aun así verosímil. Se trata, ahora sí, del comportamiento moral. Los defensores de la teoría del gen "egoísta" asumen, implícita o explícitamente, que la moralidad supone alguna forma de altruismo psicológico, es decir, la intención de beneficiar a otros individuos sin pensar en ganancias ulteriores, y especialmente a individuos fuera del estrecho círculo familiar. En esto concuerdan con los filósofos morales, probablemente la mayoría, que se oponen a la idea hobbesiana de que la moral se puede reducir a un cálculo egoísta/prudencial en el que el altruismo psicológico es redundante. Pero justo debido al vínculo entre moralidad y altruismo psicológico, los sociobiólogos predican una oposición entre la moral y la selección natural.

Ellos asumen que el altruismo psicológico implica un altruismo biológico incompatible con la tesis del gen egoísta. Su posición en este aspecto revela una falta de reflexión sobre las diversas formas de altruismo psicológico. Entienden por éste una disposición que podemos llamar "altruismo incondicionado", porque dispone al sujeto a donar beneficios indiscriminadamente, sin considerar, por ejemplo, si el receptor es un pariente cercano o no, o si tiene un carácter tal que lo disponga a la gratitud y a la reciprocidad. Si la selección natural tuviese que elegir entre un egoísta miope y un altruista incondicionado, está claro que elegiría al egoísta miope.

Pero existe otro tipo de altruismo, al que Trivers llamó "altruismo recíproco" en su célebre ensayo (1971). El altruismo recíproco es un fenotipo conductual complejo.[3] Por un lado, implica la donación

[3] Digo "conductual" porque así lo presentó Trivers en el ensayo, al definir el comportamiento por sus efectos y no por sus motivos o causas, de acuerdo con la práctica biológica de la época. Los rasgos psicológicos que controlan esa conducta en los humanos no pertenecen propiamente al rasgo biológico, pues éste puede estar presente en muchas especies que no tienen las características de la psicología humana.

de un beneficio a un individuo no emparentado con el donante; por otro, este fenotipo asegura su aptitud en razón de que el donante es capaz de dirigir su comportamiento altruista exclusivamente a aquellos individuos que exhiben un fenotipo complementario, el de la reciprocidad. Gracias a la reciprocidad de los receptores y a la relación costo-beneficio, donde el beneficio recibido por el receptor es mucho mayor que el costo en que incurre el donante, el altruismo recíproco es un comportamiento que produce ganancias netas. Si *A* gasta 2 para beneficiar a *B* con 100 y posteriormente *B* gasta 2 para beneficiar a *A* con 100, el resultado para *A* y para *B* es que han invertido 2 para ganar 100.

Trivers argumentó admirablemente en su ensayo que la moralidad humana —un sistema psicológico que incluye motivaciones altruistas condicionadas a la reciprocidad— evolucionó con la función de controlar el "altruismo recíproco" (considerado como fenotipo puramente conductual). Trivers mostró así el valor adaptativo de la moralidad en los humanos. La moralidad del altruismo recíproco por lo general ha sido entendida como una forma de hobbesianismo. Dedicaré la sección 4 de este ensayo a mostrar que se trata de una interpretación errada: el altruismo recíproco contiene un altruismo psicológico genuino, aunque condicionado.

Dando por bueno este último punto, la tesis de la oposición entre la moral y la evolución por selección natural no se apoya en la constatación de algo obvio por observación, ni tampoco es una derivación obvia de una teoría plausible y que aquí asumimos como verdadera: que los genes están obligados a codificar comportamientos que favorezcan su reproducción —comportamientos "egoístas"— so pena de desaparecer. La tesis de la oposición entre moral y selección natural se apoya más bien en una comprensión equivocada del altruismo asociado a la moral, que lo concibe como altruismo incondicionado, un fenotipo que no puede ser favorecido por la selección natural que opera sobre los genes, salvo en los casos en que el altruismo se dirige exclusivamente a individuos muy similares desde un punto de vista genético —como es el caso de los comportamientos de autosacrificio en los insectos sociales—. Alternativamente, el altruismo incondicionado entre individuos no emparentados podría haber evolucionado por selección de grupos, aunque en este caso tendría que enfrentar la fuerza contraria de selección de individuos en el interior del grupo, fuerza que, por lo general, ejerce mayor presión y opera con mayor rapidez.

La primera posibilidad para la evolución del altruismo no es relevante para el caso de la moral humana, cuyo contexto genético es muy

distinto del de los insectos sociales; y la segunda exige condiciones muy especiales y suele ser rechazada como fuerza capaz de vencer a la selección individual. Así, los biólogos evolutivos se han visto obligados a apelar a la teoría de la moral como accidente evolutivo, o a la evolución memética o cultural por cuanto que es independiente de la evolución genética (Williams 1995, pp. 340–341).

Pero su posición se basa, como vimos, en una interpretación poco verosímil del altruismo contenido en la moralidad. Es sintomático que Richard Alexander, quien también ha abordado la moral desde la perspectiva del gen egoísta (Alexander 1987), haya defendido una concepción hobbesiana de la moral, en la que ésta se deriva de un cálculo egoísta. Su propuesta parece una auténtica explicación de la moral como adaptación, pero en realidad no es sino la otra cara de la misma moneda calvinista que expresa la oposición entre altruismo genuino y naturaleza. La visión sociobiológica se resume en esta fórmula: si la selección natural explica la actitud moral, el altruismo no es parte de ella; si el altruismo es real, es un accidente y no un producto de la selección natural.

3. *La moralidad como accidente evolutivo*

Detengámonos un momento en la teoría de la moral como accidente evolutivo, que gozó de simpatizantes entre los filósofos que primero reaccionaron ante la biología evolutiva del altruismo (*cfr.* McGinn 1979 y Singer 1981). McGinn 1979, por ejemplo, construyó la oposición entre moralidad y selección natural con un concepto de moralidad que involucra esencialmente un altruismo incondicionado, el cual implica que los intereses ajenos se ponen por encima de los míos: estoy dispuesto a ceder un beneficio aunque en el receptor no exista una disposición recíproca. Si la moralidad se equipara a este tipo de altruismo, es en efecto imposible que evolucione por selección natural. El altruismo incondicionado existe en casos extraordinarios como el de la madre Teresa, pero esos casos no pueden tomarse como paradigmáticos de conducta moral. Para entender la moral es más importante la actitud imparcial, es decir, la que da el mismo peso a intereses equivalentes aunque sean de personas distintas. Esto es fundamental en un contexto cooperativo, que, sin duda, es el contexto en el cual se desenvuelve la moral.

Ahora bien, quizás sea posible defender una teoría en la que el altruismo incondicionado evolucione como efecto secundario de otro rasgo que habría sido seleccionado directamente. Toda posición de este estilo debe identificar los blancos directos de la selección natural

y explicar cómo el altruismo incondicionado, a pesar de su naturaleza biológicamente perjudicial, es seleccionado con ellos como un efecto secundario. McGinn y Singer coinciden en proponer que el altruismo es un producto derivado de la capacidad para el conocimiento objetivo de la naturaleza; esta capacidad conduce forzosamente a reconocer que otras personas también tienen intereses.

Pero reconocer la objetividad *fáctica* de los intereses de otras personas no es lo mismo que reconocer su objetividad *práctica* universal. La diferencia entre ambos es la diferencia entre pretender la verdad de "Para Julia es importante que *p* ocurra" y pretender la verdad de "Es importante que *p* ocurra". La diferencia es obvia si reparamos en que la capacidad de juzgar objetivamente sobre los intereses de Julia (juicios del primer tipo) es indispensable cuando mi intención es frustrarlos con el fin de realizar los míos. Así, la objetividad teórica en la atribución de intereses a otras personas no implica la actitud imparcial que los convierte en razones para mí. Esta segunda objetividad, y no la primera, es la que se requiere para dar el salto a la perspectiva moral.

Influyentes filósofos morales, como Bernard Williams (1985, pp. 65–69), han objetado legítimamente intentos semejantes de probar la objetividad del principio de imparcialidad apoyándose en la objetividad del conocimiento. Más aún, si pudiésemos probar la validez práctica universal de los intereses particulares, ello no implicaría altruismo incondicionado, pues todavía queda por establecer cuál es el peso relativo que se otorga a esos intereses, comparado con el que se les otorga a los propios (Sober y Wilson 1998, pp. 242–248). Si el peso relativo tiende a la igualdad, entonces estamos más bien en la esfera de la imparcialidad y no en la del altruismo incondicionado.

La teoría del altruismo incondicionado como accidente evolutivo podría reaparecer en una versión alternativa, que propusiese otro rasgo directamente seleccionado sobre el cual el altruismo incondicionado habría viajado gratis en la evolución. Pero si concebimos la moral basada en la reciprocidad y aceptamos, con Trivers, su valor adaptativo, tenemos pocos incentivos para pensar en alternativas. Como altruismo recíproco en lugar de incondicionado, la moralidad pudo haber sido un blanco directo de la selección natural. No hay, *a priori*, ninguna razón para que el egoísmo de los genes se exprese sólo en el egoísmo psicológico de los individuos. Si acaso hay algo que puede decirse *a priori* es que existe alguna plausibilidad de que lo contrario sea verdad, dado que tanto el cuidado parental como

la cooperación social se apoyan más confiablemente en motivaciones altruistas (imparciales en el segundo caso) que en motivaciones egoístas.

4. *El gen egoísta y la psicología moral hobbesiana*

No hay duda de que los defensores de la teoría del "gen egoísta", como G.C. Williams o Dawkins, tenían conocimiento de como explicó Trivers la evolución del altruismo condicionado a la reciprocidad. ¿Cómo se explica, entonces, que hayan insistido en su tesis de la oposición entre moral y selección natural?

La respuesta exige reconocer una sutil ambigüedad en la idea de altruismo recíproco. La condición de la reciprocidad puede entenderse como la subordinación consciente de *actos* altruistas a beneficios ulteriores para el agente, y, en este sentido, como un genuino egoísmo psicológico (*cfr.* la nota 2). Probablemente así vieron G.C. Williams y Dawkins el modelo de Trivers y lo consideraron insuficiente para capturar el fenómeno de la moral; sin embargo, se puede mostrar que el modelo de Trivers contiene un altruismo genuino. Trivers entendió la moralidad humana como un complejo sistema psicológico para controlar el altruismo recíproco como fenotipo conductual. Dentro de ese sistema psicológico reconoció la existencia de motivaciones genuinamente altruistas. Su ensayo "The Evolution of Reciprocal Altruism", publicado originalmente en 1971, es muy claro en este punto: afirmó que la selección favorecería una actitud desconfiada hacia actos altruistas generados por mero cálculo egoísta, sin la "base emocional de la generosidad". Sin esa base, el altruista no es consistente a través del tiempo (Trivers 1971, pp. 50–51). La actitud desconfiada estaría basada en mecanismos psicológicos que detectan los verdaderos motivos del comportamiento y funcionaría como una presión selectiva contra los calculadores egoístas. Trivers encontró que los estudios psicológicos apoyaban esta hipótesis, pues

> existe amplia evidencia a favor de la noción de que los humanos responden a los actos altruistas según los motivos que perciben en el agente. Responden con mayor altruismo cuando perciben al otro como a un altruista genuino, es decir, como alguien que ofrece un acto altruista como un fin en sí mismo, sin dirigirlo a obtener una ganancia ulterior. (Trivers 1971, p. 51)

Es obvio que Trivers concebía al altruista recíproco como psicológicamente distinto del frío calculador que dispensa beneficios pensando

solamente en las ganancias que vendrán, al menos así se nota en su ensayo de 1971. Trivers es allí partidario de un proceso de selección psicológica que favorece las motivaciones altruistas. Defiende el carácter adaptativo de los mecanismos de detección de motivaciones altruistas y de las preferencias por esas motivaciones en quienes son parte de una empresa cooperativa. Estas mismas preferencias terminarían por funcionar como agentes selectores si asumiéramos que quienes son consistentemente rechazados como partes en empresas cooperativas tienen un menor éxito reproductivo. Así, la selección favorecería finalmente a los individuos con motivaciones genuinamente altruistas en los intercambios sociales.

Williams y Dawkins, al parecer, no prestaron atención a estos pasajes de Trivers. Existen indicios textuales indirectos de que, en su manera de entender el altruismo recíproco, ellos no incluyen las motivaciones genuinamente altruistas que Trivers sí incluía en ese modelo. En la breve exposición que dedica al altruismo recíproco, G.C. Williams (1995, p. 326) se refiere a Darwin de manera aprobatoria diciendo que ya él había llamado a la reciprocidad un "motivo mezquino" [*a low motive*].[4] Por su parte, en el capítulo 10 de *El gen egoísta*, Dawkins expone el altruismo recíproco de Trivers con ayuda del concepto de *estrategia evolutivamente estable* de J. Maynard Smith. Y si bien allí no califica en absoluto los motivos de la estrategia del altruismo recíproco, sí le pone el nombre sintomático de "rencorosa" (*Grudger*). Ello se debe a que el altruista recíproco castiga al interactor que no coopera con un acto de no cooperación en la siguiente jugada.

Sin duda, la idea de reciprocidad admite una ambigüedad que la hace vulnerable a interpretaciones basadas en el egoísmo motivacional. Así se explica que el concepto de altruismo recíproco haya dado lugar a los desarrollos hobbesianos del biólogo Richard Alexander, para quien el motivo real de la reciprocidad es un cálculo egoísta de largo plazo, acompañado del deseo de establecer para sí mismo una reputación como interactor confiable. Todavía hoy vemos a los teóricos de la evolución del altruismo refiriéndose a Trivers (1971) como un ejemplo de hobbesianismo:

[4] G. Williams tiene en mente el siguiente pasaje: "a medida que mejoraran los poderes de inferencia y de previsión de los miembros [de una tribu], pronto cada uno aprendería que, como regla general, si ayudaba a sus prójimos, recibiría ayuda a cambio. Este motivo mezquino lo llevaría a adquirir el hábito de ayudar a sus compañeros", Darwin 1989/1877, pp. 130 y s.

> La fuerza explicativa de la teoría de la aptitud inclusiva y del altruismo recíproco (Hamilton 1964; Trivers 1971; Williams 1966) convenció a una generación de investigadores de que lo que parece ser altruismo —sacrificio personal en beneficio de otros— en realidad no es sino egoísmo a largo plazo. (Gintis *et al.* 2003, pp. 153–154)

Por las razones ya explicadas, me parece que esta interpretación está equivocada. Quienes creen que el altruismo genuino no existe y que la moral es sólo un cálculo egoísta, mezclado con la apariencia engañosa y autoengañosa de altruismo, fuerzan el modelo de Trivers dentro de ese molde. Esto se aplica al ya mencionado Alexander, quien ha defendido que nos engañamos a nosotros mismos creyéndonos altruistas (Alexander 1987, p. 123), para ocultar mejor a los demás el deseo de ganancia que secretamente anima a todas nuestras interacciones sociales.

Quizás la psicología moral humana sea como Hobbes y los hobbesianos la describen; pero no creo que la teoría de la evolución por selección natural nos dé evidencias decisivas de que así es. No hay nada en la idea de la moral como una adaptación biológica que nos obligue a adoptar una versión hobbesiana de la moral: una forma sofisticada de egoísmo que consista en enfocarse en los beneficios personales a muy largo plazo. Contra ello están las consideraciones evolucionistas sobre el cuidado parental y sobre el altruismo recíproco en Trivers. La explicación adaptacionista del surgimiento de motivaciones altruistas respecto de extraños al estilo de Trivers cuadra perfectamente dentro de la teoría del gen egoísta. Existen, en cambio, razones para dudar de los argumentos de los sociobiólogos, cuando los dan, para oponer la selección natural a la posibilidad del altruismo. En otro lugar he defendido que la explicación de la valoración positiva de las motivaciones altruistas como una forma de autoengaño favorecido por la selección natural (Alexander 1987) es inconsistente (Rosas 2004). Si la psicología humana no contiene un genuino altruismo por diseño evolutivo, la razón no es una presunta oposición de principio entre la selección natural y las motivaciones genuinamente altruistas, incluso si aceptamos que el gen es la unidad de selección.

Lo contrario me sigue pareciendo más verosímil; pero creo que es importante no caer en el juego de querer decidir esta cuestión *a priori* a partir del concepto de selección natural. Esta estrategia es, en realidad, tan poco creíble como la estrategia filosófica tradicional de resolver con argumentos *a priori* si las motivaciones humanas son necesariamente egoístas o admiten un genuino interés por los demás. Este debate sólo puede resolverse recurriendo a evidencias de tipo

experimental. Sin entrar aquí en detalles, vale la pena mencionar brevemente dos líneas de investigación experimental que arrojan un balance positivo en favor de la existencia de motivaciones altruistas.

Daniel Batson y sus colaboradores han llevado a cabo una serie de experimentos diseñados para refutar las distintas versiones del egoísmo psicológico, a saber, la teoría que dice que toda acción humana se hace con vistas a un beneficio que obtiene el que actúa (*cfr.* Batson y Shaw 1991). La teoría implica que si ayudamos a otro, eso ocurre siempre como medio para obtener algún beneficio personal, que es el fin último de toda acción. El programa experimental de Batson y sus colegas busca desprestigiar el egoísmo psicológico, entendido como un programa de investigación que pone por hipótesis que *algún tipo* de beneficio egoísta subyace a toda acción humana, incluso de las acciones encaminadas a ayudar. Como programa, es más difícil de refutar que una teoría que postule *un solo tipo particular* de beneficio personal. Si se refuta una propuesta del beneficio personal como fin de una acción de ayudar, siempre es posible pensar en otro tipo de beneficio, distinto del anterior, como explicación de esa misma acción. Batson y sus colegas diseñaron un tipo de experimento para refutar, una tras otra, las distintas hipótesis egoístas y su éxito ha desprestigiado, en buena medida, al egoísmo psicológico.

Más importante para el altruismo como equidad es otra línea de investigación proveniente de la economía experimental. Desde la década de 1980, una serie de experimentos ha puesto en duda el supuesto canónico de la economía, según el cual el agente económico es totalmente interesado. Los experimentos enfrentan a los sujetos en juegos como el "ultimátum" o el "dictador", y el comportamiento observado muestra que los sujetos no buscan simplemente la mayor ganancia. De manera consistente se desvían de la predicción derivada del supuesto canónico del agente guiado exclusivamente por su propio interés. El balance de estos experimentos es que los humanos nos preocupamos por la equidad en las interacciones sociales. Los seres humanos reconocemos la validez de un punto de vista imparcial, desde el cual los demás valen tanto como nosotros mismos (Fehr y Gächter 2002; Fehr y Fischbacher 2003).

5. *¿Adquisición cultural o adaptación ancestral?*

Si esas líneas de investigación experimental tienen éxito y demuestran la realidad del altruismo psicológico, no por ello quedaría establecido que el comportamiento moral es producto de la selección natural, pues podría ser que esa motivación fuera *adaptativa*, es decir, que

se ajuste a las necesidades de los individuos que la tienen sin ser una *adaptación*, un producto de la evolución por selección natural. Es propio de los organismos inteligentes aprender conductas o disposiciones que se adaptan a sus necesidades en las condiciones cambiantes del entorno. Las motivaciones altruistas de respeto mutuo pueden ser disposiciones aprendidas, inculcadas por educación; fabricaciones culturales muy exitosas, en la medida en que ayudan al bienestar de sus portadores.

Ahora bien, el hecho de que toda cultura humana conocida tenga normas de respeto mutuo ¿no sugiere acaso que estamos frente a algo diferente de una adquisición cultural? ¿No es esta universalidad signo de un carácter innato? No necesariamente. Podría tratarse de una respuesta aprendida que permitió a los grupos humanos ajustarse a condiciones históricamente contingentes, pero que han perdurado durante los últimos veinte mil años, un lapso quizás breve para que hablemos de adaptaciones complejas. Su utilidad para las condiciones de vida de los humanos modernos pudo haber llevado a su difusión por todas las poblaciones humanas, o incluso a que las diferentes culturas las hubiesen inventado independientemente.

La estrategia más contundente contra estas dudas es acudir directamente a nuestro pasado evolutivo a buscar evidencias de disposiciones semejantes. Esto no es fácil, pues reconstruir las conductas y el modo de vida de nuestros ancestros con base en fósiles y en consideraciones de ecología comparada parece una oportunidad para la fantasía especulativa de cada cual. ¿Existe algún modo de constreñir este ejercicio dentro de parámetros empíricos bien establecidos? Creo que un estudio serio de la psicología de los primates ofrece un camino digno de ser explorado. Supongamos que la moral apareció temprano en la filogenia que condujo hasta los humanos modernos, quizás en el género de los australopitecos (el primer homínido que fue bípedo con certeza). Si es así, debían haber existido versiones de ese rasgo en el ancestro común, del cual hace aproximadamente seis millones de años se separaron las ramas que llevan al chimpancé y a los humanos modernos. Estas versiones presumiblemente se habrían heredado a la rama de los chimpancés, donde hoy podríamos observarlas. Si las versiones fuesen mucho más ancestrales, las encontraríamos también en primates pertenecientes a ramas filogenéticas más alejadas.

6. *La analogía con los primates y las piezas de la moralidad*

La analogía con los primates nos permite acceder al pasado de nuestra especie a partir de la comparación con especies filogenéticamente

próximas. El procedimiento es identificar rasgos que un ancestro común hereda a sus descendientes, que luego se bifurcan en especies distintas, como es el caso de los chimpancés y el linaje homínido que conduce a los humanos modernos. Frans de Waal, primatólogo de Emory University y director de Living Links, ha dedicado varios años a buscar, en nuestros parientes primates más cercanos, evidencias empíricas de comportamientos y motivaciones que puedan identificarse como morales. En publicaciones en las que ha presentado y resumido sus ideas sobre este tema (*cfr.* en especial de Waal 1996, y Flack y de Waal 2000), defiende que hay evidencia empírica para atribuir cuatro tipos de capacidades psicológicas que sugieren una moral rudimentaria en nuestros parientes cercanos: capacidades para la simpatía, para la reciprocidad, para la representación de reglas de conducta y para la solución de conflictos que amenazan la estabilidad del grupo. Estas mismas capacidades debían estar ya presentes en nuestro primer ancestro homínido.

Su proyecto consiste en establecer una alternativa radical a la que ofrecen quienes defienden, de una u otra forma, la oposición entre naturaleza y moralidad. El filósofo Daniel Dennett, por ejemplo, ha dicho que los chimpancés se muestran en su vida social como "verdaderos habitantes del estado de naturaleza hobbesiano, mucho más crueles y brutales de lo que muchos quisieran creer" (Dennett 1995, p. 481).[5] Aunque esta opinión es compatible con una defensa de la moralidad como rasgo derivado en el linaje homínido, de Waal cree que ella abre una puerta a la tesis calvinista. Sostiene que la evidencia empírica apoya más bien la existencia de "piezas constitutivas" [*building blocks*] de la moral en especies de primates como los monos capuchinos, los chimpancés, los bonobos y los macacos, entre otros.

La formulación de su hipótesis en términos de evidencias de "piezas constitutivas", o "prerrequisitos" de la moral (*cfr.* Flack y de Waal 2000, *passim* y p. 3), contiene una ambigüedad que propició una crítica aguda y justificada. Un componente de la moralidad no es necesariamente un componente moral (Thierry 2000). El comportamiento moral se apoya en una arquitectura psicológica compleja, que incluye un conjunto de variados mecanismos psicológicos. El hecho de que el conjunto completo sea suficiente para la moral no significa que un subconjunto cualquiera lo sea. La interpretación que de Waal emprende del comportamiento de los primates no humanos

[5] A propósito de la brutalidad en chimpancés y otros primates, *cfr.* Wrangham y Peterson 1996, donde se presenta evidencia sólida para atribuir la brutalidad sólo a los machos, como su rasgo característico.

está encaminada a mostrar en ellos la presencia de varios mecanismos psicológicos *necesarios* en cualquier sistema moral. Pero en ausencia de una tesis que defienda la *suficiencia* de esos mecanismos, su presencia en algunas especies emparentadas con nosotros no garantiza que sus miembros, o grupos de ellos, puedan clasificarse como agentes o comunidades morales.

Esto puede ilustrarse retomando la observación de Dennett. Un representante del estado hobbessiano de naturaleza tiene que tener capacidades para la prudencia y para leer las intenciones de los individuos con los que compite. Emplea estas capacidades para salir airoso en confrontaciones eminentemente competitivas, en las que no hay rastro alguno de constreñimientos morales. Supongamos que con base en estas capacidades y por razones puramente instrumentales,[6] algunos individuos en este estado natural se viesen motivados a cumplir las leyes naturales dirigidas a la obtención y el mantenimiento de la paz, como Hobbes propone. Hobbes los consideraría ya por eso agentes morales, porque él entiende que las reglas morales son instrumentales y nada más. Asumo aquí que ésa no es la posición filosófica más convincente sobre la moral y que los agentes en cuestión no califican como agentes morales si sólo siguen reglas morales en un sentido instrumental. Sin embargo, las capacidades que los caracterizan son plausiblemente capacidades que un verdadero agente moral también debería tener (son necesarias, pero no suficientes). Son, pues, prerrequisitos o piezas constitutivas en sentido débil.

De Waal concede que la presencia de uno o varios prerrequisitos en el sentido débil no hacen de un organismo un agente moral; y probablemente concedería que atribuir a los primates no humanos el carácter de agentes hobbesianos es una tesis suficientemente interesante en el contexto de nuestros conocimientos actuales, pero su proyecto es distinto. Él cree que, tomados en conjunto, los mecanismos cuya presencia defiende en algunos primates no humanos hacen muy plausible una tesis más fuerte. La tesis es que algunos primates no humanos —y los chimpancés parecen ser candidatos promisorios— presentan estados mentales o motivaciones que son característicamente morales. Al menos en tres ocasiones interpreta la evidencia empírica en este sentido. Refiriéndose a las intervenciones del macho alfa en los conflictos entre terceros, de Waal afirma que son imparciales y agrega: "La habilidad para poner así las propias prefe-

[6] "Instrumental" quiere decir aquí que se adoptan reglas morales o se persigue la paz únicamente como un medio que se juzga adecuado para el fin de la propia conservación o bienestar.

rencias a un lado es otro indicio de que una forma rudimentaria de justicia puede existir en los sistemas sociales de los primates no humanos" (Flack y de Waal 2000, p. 12). Hablando también de los chimpancés, de Waal menciona patrones de comportamiento que, en su opinión, indican que se preocupan por la comunidad, que tienen una capacidad para "reconocer el valor de una coexistencia armoniosa para alcanzar intereses compartidos [...], incluso si ello requiere, al menos en ocasiones, la subordinación de intereses privados (no compartidos) a intereses comunitarios [...]" (Flack y de Waal 2000, p. 15).

En otro pasaje, hablando de los mecanismos o "piezas constitutivas" de la moral presentes en primates, de Waal señala que estas piezas

> son fundamentales para los sistemas morales, porque ayudan a generar conexiones entre individuos en sociedades humanas o animales a pesar de los conflictos de intereses que inevitablemente surgen. [...] Estos mecanismos facilitan la interacción social cooperativa porque exigen a los individuos hacer "compromisos" en sus cursos de acción que más tarde pueden resultar contrarios a los intereses individuales [...]. (Flack y de Waal 2000, p. 3)

En estos tres textos se atribuye a los chimpancés la capacidad de postergar la satisfacción de un interés individual en aras de lograr un interés común. Aunque es posible hacer una lectura no moral de esta capacidad, los textos sugieren una lectura moral. De Waal quiere que su hipótesis de las "piezas constitutivas" se entienda en sentido fuerte y toma las observaciones que la apoyan como evidencia de la capacidad de poner los intereses comunes por encima de los individuales.

Así las cosas, su propuesta debe evaluarse por la pertinencia de sus interpretaciones psicológicas de las evidencias conductuales aportadas por diferentes estudios. Sus críticos le han objetado que tiñe sus descripciones de la evidencia conductual con connotaciones morales sin apoyo real en evidencia empírica (*cfr.* Kummer 2000; Thierry 2000). Señalan, por ejemplo, que la tesis de la presencia de un "interés por la comunidad" en los chimpancés requeriría un concepto de comunidad cuya presencia no se demuestra. De igual modo, niegan que haya verdadera evidencia de imparcialidad en las intervenciones de los machos en conflictos entre terceros.

El propio de Waal es consciente de que aún hay espacio para alternativas de interpretación y para la búsqueda de nuevos datos. En

lo que sigue cuestionaré sus interpretaciones desde el punto de vista de las capacidades cognitivas implícitamente atribuidas a los primates no humanos; al mismo tiempo ofreceré alternativas de interpretación que expliquen la conducta observada.

7. *Alternativas de interpretación*

7.1. Reciprocidad

Las interpretaciones de la conducta de los primates propuestas por de Waal tienen como denominador común la atribución de capacidades cognitivas relativamente sofisticadas. En el caso de las capacidades para la reciprocidad, de Waal contrapone dos lecturas: reciprocidad "calculada" y reciprocidad "simétrica", basada en la asociación. En *Good Natured* (1996, pp. 157 s.), de Waal explicaba la diferencia entre ambas de este modo: la reciprocidad simétrica es producto accidental de la asociación, y el individuo no guarda un registro de la historia de los intercambios; la reciprocidad calculada, en cambio, exige guardar un registro en la memoria y por eso es más exigente, cognitivamente hablando.

Stevens y Hauser (2004) han defendido recientemente que la reciprocidad en los animales está seriamente constreñida por la memoria. La asociación estable entre individuos, al reducir el número de interactores que hay que recordar, reduce también las exigencias sobre esta capacidad. "Guardar un registro de las obligaciones de reciprocidad con múltiples individuos puede representar una tarea computacionalmente intensiva para la memoria" (Stevens y Hauser 2004, p. 64).[7] La asociación facilita así el intercambio recíproco; basta recordar la última actuación del asociado y marcarlo positiva o negativamente para continuar o suspender la asociación, según sea el caso.

La capacidad de memoria y de aprendizaje también constriñe la *complejidad* de la estrategia utilizada. Poniendo como ejemplos las estrategias *Pavlov* (ganar-quedarse/perder-irse) y TFT, Stevens y Hauser señalan que los humanos generalmente utilizan *Pavlov* en experimentos que les permiten aplicar toda su memoria en el juego del dilema del prisionero, pero que emplean TFT si su memoria está recargada con tareas adicionales (Stevens y Hauser 2004, p. 64) Esto sugiere que TFT con asociados es la estrategia más probable en los

[7] "Keeping score of reciprocal obligations with multiple individuals can be a computationally intensive burden on memory."

animales, pues sólo exige recordar cómo se marcó al asociado con base en su última actuación.

Pero las exigencias sobre la memoria no constituyen el factor más importante para distinguir distintos tipos de reciprocidad. Más importante es que de Waal concibe la "reciprocidad calculada" basada en la capacidad de seguir o acatar reglas que se saben compartidas. Seguir reglas sociales es una capacidad cognitiva muy sofisticada que requiere la representación de normas prescriptivas de conducta. En *Good Natured*, esta tesis se apoya en otra que defiende que los chimpancés y otros primates tienen expectativas sobre el comportamiento de los demás. El argumento es éste: si un animal tiene expectativas sobre lo que sucederá, la explicación es que puede representarse un curso regular de acontecimientos. Pero las expectativas no son algo que podamos observar directamente; de Waal da entonces un criterio conductual de la presencia de expectativas: un animal tiene la expectativa de un acontecimiento si la ausencia de éste le produce confusión, sorpresa y ansiedad, y eso es lo que se observa en algunos experimentos controlados (de Waal 1996, p. 96).

Asumiendo la transparencia conductual de estados mentales como la confusión o la sorpresa, se puede objetar, no obstante, que la presencia de expectativas revela que el animal se representa regularidades, pero no revela si distingue entre regularidades causales y regularidades que dependen del acatamiento de normas o reglas sociales (de Waal 1996, pp. 90, 95). Nótese que si *A* espera un comportamiento regular en *B*, esto no significa que se representa a *B* como capaz de actuar por la conciencia de una regla. Todo en la naturaleza actúa de acuerdo con leyes, pero sólo los seres racionales actúan por representación de las leyes, dijo un célebre filósofo.[8] Y así como la primera forma de seguir una regla no implica la segunda, tampoco la capacidad de representarse un proceso regular en el primer sentido implica la capacidad de representarse acciones que siguen reglas en el segundo. No basta, entonces, que un animal muestre los signos conductuales de la expectativa para concluir que se representa el comportamiento de otro como si estuviera sujeto a una obligación.

Ingeniosamente, de Waal sostiene que se puede inferir la representación de obligaciones, a diferencia de simples regularidades naturales, cuando los signos conductuales de expectativas defraudadas (confusión, sorpresa) se presentan como comportamiento agresivo (castigo). Según de Waal, los chimpancés y los monos capuchinos,

[8] I. Kant, *Fundamentación de la metafísica de las costumbres*, segunda sección.

que se caracterizan por compartir comida sin restricciones de parentesco ni de rango (1996, pp. 144, 148), tienen expectativas de reciprocidad, razón por la cual reaccionan con "agresión moralista" si estas expectativas son defraudadas o ignoradas (1996, pp. 97, 159–160, 189–190). Nótese que si se acepta esta interpretación del comportamiento agresivo, habría que aceptar que los chimpancés son capaces de representarse a otros individuos como mentes capaces de representarse el peso de los imperativos de conducta. Esto supone que los primates tienen la capacidad de atribuir estados mentales, un punto fundamental para la tesis de la moralidad en los primates, y sobre el que volveré más adelante.

Esta lectura de la agresividad es sugestiva, pero hay una explicación cognitivamente menos exigente del comportamiento recíproco observado. Para explicar la reciprocidad, basta postular en *A* la disposición a comportarse amistosa/agresivamente con *B*, si *B* está marcado positiva/negativamente en la memoria de *A*. La marca positiva/negativa depende del carácter amistoso/agresivo de las actitudes anteriores de *B* hacia *A*. Esto quiere decir que las disposiciones de *A* hacia *B* son *causadas* por la representación positiva/negativa que *A* tiene de *B*, pero ello no implica que *A* adquiera esas disposiciones porque se representa su interacción con *B* sujeta a obligaciones sociales compartidas. Las disposiciones *ocurren* en *A* causadas por una representación de *B* en su memoria, pero la representación de una obligación socialmente "compartida" no interviene como factor causal del comportamiento de *A*.

7.2. Interés por la comunidad

Las capacidades cognitivas que de Waal atribuye a los chimpancés alcanzan un nivel similar de sofisticación en los textos citados en la sección 6. Allí afirma que los chimpancés parecen "reconocer el valor" de una existencia armoniosa, que "ponen a un lado" sus propias preferencias y que se "comprometen" con cursos de acción que resultan contrarios a sus intereses individuales. Se sugiere así que deliberan y anteponen *conscientemente* la satisfacción de los intereses del grupo a la satisfacción de sus intereses inmediatos.

Existen dos posibles interpretaciones de esa atribución. Si el individuo decide satisfacer los intereses grupales por juzgar que el grupo es el medio indispensable para satisfacer sus intereses individuales a largo plazo, se trata de una deliberación prudencial. En cambio, si el individuo cede a los intereses del grupo por considerar que intereses semejantes merecen igual consideración, sin importar de

quién son esos intereses, el aplazamiento de la satisfacción particular tiene el carácter que distinguimos como moral. Cualquiera de estas dos interpretaciones atribuye a los chimpancés capacidades cognitivas sofisticadas. La diferencia entre ambas es más un asunto de la motivación operante, a saber: si se tiene o no la capacidad de atender a los intereses de los otros miembros del grupo como fines valiosos en sí mismos. De Waal escoge la primera interpretación, lo cual plantea a su proyecto problemas que más adelante explicitaré.[9]

No estoy en contra de atribuir a los chimpancés capacidades representativas o cognitivas; sin embargo, tener un concepto del grupo y de los individuos que lo componen como sujetos de intereses con los que mis propios intereses coinciden a largo plazo, aunque pueden chocar en lo inmediato, exige capacidades cognitivas complejas. La oposición entre el interés inmediato y el de largo plazo sólo se percibe si el agente es capaz de representarse los efectos inmediatos y lejanos de su acción y el modo en que inciden sobre su bienestar como sujeto que se extiende en el tiempo. Esto no podría hacerse sin proyectarse contrafácticamente hacia el futuro.[10] Parece más verosímil explicar los comportamientos observados con base en intereses inmediatos derivados del curso regular de su vida social, en lugar de atribuirles la conciencia de que la satisfacción de intereses inmediatos puede implicar la frustración de otros más duraderos.

Las intervenciones "imparciales" de los machos que ejercen el papel de pacificadores (Flack y de Waal 2000, p. 12) se podrían explicar por una intención diferente del deseo de ser imparcial para preservar la estabilidad del grupo. Quizás el macho alfa tema que una intervención parcial tenga por efecto el fortalecimiento de algún individuo o

[9] Según de Waal, el chimpancé no sacrifica sus intereses individuales a los del grupo; más bien, comprende que la satisfacción de intereses comunes es indispensable para lograr la satisfacción de los intereses individuales (1996, p. 31, y Flack y de Waal 2000, pp. 14–15). Aunque de Waal no parece advertirlo, esto acerca su comprensión de los chimpancés a la apreciación explícitamente rechazada por él, según la cual los chimpancés serían "verdaderos representantes del estado natural hobbesiano" (*cfr.* Dennett 1995, p. 483).

[10] El compendio más autorizado sobre cognición en primates (*cfr.* Tomasello y Call 1997) no dedica ningún capítulo ni sección a la comprensión del tiempo; la palabra 'tiempo' o sus derivados tampoco aparece en el índice de conceptos. La renuencia a admitir que los chimpancés puedan proyectarse contrafácticamente hacia el futuro se basa en evidencias indirectas; por ejemplo, que los chimpancés no parecen representarse ni referirse a entidades meramente posibles, es decir, entidades de las que no hay ejemplares en el entorno inmediato, lo cual se revela en su carencia de cooperación orientada hacia futuros posibles como la que se observa en los humanos (*cfr.* Brinck y Gärdenfors 2003).

de una coalición que esté en busca de la posición alfa. La intervención imparcial, en cambio, impide ese fortalecimiento. La intención sería aquí semejante a la que parece animar otro comportamiento similar observado por de Waal, cuando el alfa interviene para impedir la formación de amistades o coaliciones entre machos adultos. En la presunta intervención imparcial, el macho alfa quiere impedir que haya ganadores en una disputa.

Una lectura semejante podría aplicarse a otras intervenciones pacificadoras mencionadas por de Waal. Con ello no se evita la atribución de intenciones y de la capacidad de entender y predecir el comportamiento de los demás miembros del grupo, pero se limita la atribución de intenciones a aquellas que representan beneficios inmediatos para el individuo que interviene. No se postula una capacidad compleja de deliberación, en la que el individuo se proyecta contrafácticamente hacia el futuro, comprendiendo la articulación entre beneficios a largo y a corto plazo. Se evita así suponer que el individuo pacificador tiene en mente la vida armónica del grupo como condición del propio bienestar a largo plazo.[11]

8. *Las piezas de la moralidad: la versión débil*

El constreñimiento de los intereses individuales admite, como vimos, dos interpretaciones posibles. De Waal prefiere atribuir a los chimpancés la interpretación hobbesiana o instrumental del constreñimiento. Dentro de la concepción dominante —no hobbesiana— en la filosofía moral, esta interpretación implica negar capacidades genuinamente morales a los chimpancés, aunque sin duda les atribuye una psicología sofisticada con elementos necesarios, pero no suficientes para una auténtica capacidad moral.

Esto vuelve a poner sobre el tapete el sentido de la hipótesis de las "piezas constitutivas", pues la interpretación hobbesiana es la que

[11] Sea cual sea la intención de los individuos, las conductas pacificadoras o conciliadoras pueden tener como *función* controlar y reducir el nivel de agresión en el interior del grupo y mantener así su estabilidad, en especial si hay razones para pensar que la capacidad de vivir en grupo fue importante para su supervivencia en el pasado. Pero no se debe confundir la *función* de su comportamiento con la *intención* del mismo. De Waal está perfectamente consciente de este punto (Flack y de Waal 2000, p. 74). Una cosa es decir que un comportamiento cumple la función de mantener la estabilidad del grupo, y otra muy distinta es decir que el agente de esa conducta haya tenido expresamente la intención de producir precisamente ese efecto. El comportamiento puede cumplir esa función incluso siendo efecto de un mecanismo psicológico en el que no intervienen intenciones. En los chimpancés me parece verosímil atribuir intenciones, pero éstas no tienen por qué incluir una representación del efecto por el cual el comportamiento fue seleccionado.

pusimos como ejemplo de una interpretación débil. La interpretación débil se caracteriza por atribuir a los chimpancés capacidades psicológicas necesarias, aunque no suficientes, para la moralidad. Los agentes hobbesianos son una versión posible de la interpretación débil; tienen como fin último su interés individual y adoptan reglas morales, si acaso, sólo instrumentalmente. Sin embargo, ya vimos que existen razones de peso para poner en duda la atribución de este tipo de racionalidad práctica incluso a los simios más inteligentes.

En última instancia, de Waal quiere defender que la moralidad es una adaptación y que no se opone a la selección natural. La interpretación débil de las "piezas constitutivas" es interesante porque permite salvar esta intención original sin necesidad de llegar a defender la existencia de rasgos morales en los primates no humanos. Para salvar esta tesis, basta la representación de una progresión evolutiva que nos lleve de las capacidades que observamos en los primates a la moralidad humana. Los agentes hobbesianos son una estación posible en este continuo; la apreciación de los chimpancés como verdaderos habitantes del estado natural hobbesiano es, por tanto, compatible con la tesis de la moral como producto de la selección natural. La moralidad pudo haber evolucionado en el linaje homínido a partir de una estación hobbesiana y, llevando este razonamiento al extremo, se puede postular también un estadio previo al de un agente hobbesiano, del cual la selección natural haya partido para diseñar, pasando por agentes hobbesianos, la moralidad como adaptación biológica.

El estado previo al agente hobbesiano estaría ejemplificado por los chimpancés tal como los he presentado aquí, y probablemente por el ancestro común de chimpancés y humanos modernos. En este estadio, el agente posee la capacidad de razonar instrumentalmente en contextos sociales, decidiendo acciones como medios para beneficios inmediatos;[12] mantiene asociaciones con pocos individuos a la vez e

[12] La conducta instrumental se ha comprobado experimentalmente en especies animales pertenecientes a diversos taxa, pero los ejemplos más conocidos son las ratas de Anthony Dickinson. El conductismo explicaba la conducta instrumental [*operant conditioning*] sin atribuir a los sujetos ningún tipo de actividad cognitiva; sin embargo, la revolución cognitiva en psicología ha abierto nuevas perspectivas de interpretación. La evidencia experimental permite postular representaciones mentales en los animales con conducta instrumental e incluso permite atribuirles capacidad para manipular representaciones de manera lógicamente consistente. Estos razonamientos instrumentales consistirían en la manipulación protológica de representaciones que en la mente animal ocupan el lugar de acciones y circunstancias, precisamente con el fin de adaptar acciones a circunstancias novedosas. Una discusión filosóficamente relevante de esta nueva psicología animal puede leerse en Papineau 2003 y en Hurley 2003.

intercambia beneficios con ellos por medio de un mecanismo que lo predispone amistosamente siempre y cuando el asociado esté positivamente marcado en su memoria. Esta marca positiva depende de haber recibido, en la última interacción, alguno de los beneficios que son usuales en su vida social, y no se basa en la conciencia de reglas que prescriban la reciprocidad. La ausencia de esta conciencia se explica por una capacidad muy limitada de entender el comportamiento de otros agentes como sujetos de estados mentales;[13] ésta podría limitarse a atribuir emociones e intenciones de acción inmediata, y quizás estados rudimentarios semejantes a las creencias episódicas: estados de atención visual presente sobre objetos.

Para dar el paso hacia una adopción instrumental de reglas morales es necesario adquirir la capacidad de representarse reglas prescriptivas como tales. La capacidad de representarse estados de cosas, quizás en forma asociativa más que proposicional, no implica la capacidad de referirse a las representaciones mismas. Esta capacidad debe adquirirse antes de poder representarse reglas prescriptivas. La capacidad de representarse reglas de acción comenzó probablemente por las reglas de acción instrumental, que dependen de la representación (o metarrepresentación) de creencias sobre regularidades causales. La metarrepresentación de creencias como tales implica un concepto complejo y acabado de mente y de sujeto de estados mentales.

Una vez que los homínidos adquirieron la capacidad de representarse reglas de acción, pudieron adoptar una adhesión instrumental, y más precisamente egoísta, a las reglas que prescriben un equilibrio en la distribución de costos y beneficios entre los agentes involucrados en una interacción. La razón de esta adhesión es que una regla no equitativa evocaría una actitud negativa, quizás violenta, en los agentes desfavorecidos por la regla, poniendo en peligro la existencia de la interacción misma.

El carácter instrumental de la adhesión a reglas equitativas implica que el agente prefiere reglas no equitativas que lo favorezcan por encima de otros miembros de su grupo. Si se dan condiciones favorables para que el individuo gane explotando a los demás, éste las aprovechará. Aquí entra la reflexión de Trivers que ya resaltamos antes. Trivers asume que las condiciones para la explotación se presentaron en el pasado evolutivo de la especie humana; partiendo de ahí, su argumento a favor de la evolución de sentimientos altruistas señala que la actitud del explotador, apoyado ya en la fuerza, ya en el engaño,

[13] Una buena discusión crítica de los experimentos que buscaron indicios de esta capacidad en los chimpancés es la de Celia Heyes (1993).

habría puesto una presión selectiva para desarrollar la capacidad de evaluar a los posibles socios con base en la percepción de su carácter y sus motivos. Sólo los individuos con carácter altruista habrían sido escogidos consistentemente como socios en una empresa cooperativa y habrían gozado de los frutos de la cooperación. Este proceso evolutivo involucra varios elementos: la evolución de capacidades de detección de las disposiciones de carácter, pero también expresiones más o menos emocionales y automáticas de esas disposiciones, así como también expresiones inequívocas de la disposición a reaccionar adaptativamente a las disposiciones detectadas en los otros, respondiendo al altruismo con altruismo, al cálculo egoísta con desconfianza y al engaño o fraude con indignación y agresión moralista.

El desarrollo en detalle de este proceso evolutivo en sus distintos estadios y su confrontación con la evidencia empírica quedan como proyecto para el futuro. Un proyecto de este tipo se enfoca en las piezas constitutivas de los mecanismos próximos que controlan el comportamiento moral. En este sentido, no está reñido con otros proyectos que investigan la evolución de la justicia o del contrato social con la herramienta de la teoría evolutiva de juegos, aunque sea diferente de ellos (Skyrms 1996). El proyecto de Skyrms trabaja con un concepto puramente conductual del fenotipo moral. El comportamiento moral se identifica allí por sus efectos en términos de aptitud [*fitness*], no por los mecanismos próximos que lo controlan. Sus resultados son también valiosos porque muestran que la justicia paga (por lo menos a largo plazo, es decir, en términos de aptitud); pero no pueden iluminar la moralidad como fenotipo psicológico, y es como tal que ella ha interesado tradicionalmente a la filosofía.

El proyecto aquí descrito pretende iluminar la evolución de los mecanismos próximos —psicológicos— responsables del comportamiento moral. Particularmente importantes en este contexto son las investigaciones sobre la psicología de los primates no humanos. En relación con la idea promisoria de Frans de Waal de buscar allí piezas constitutivas de la moralidad, interesaba argumentar en especial que la idea puede asumirse en una forma más liberal, y acaso más plausible, que la del pionero etólogo holandés.

BIBLIOGRAFÍA

Alexander, R., 1987, *The Biology of Moral Systems*, Aldine de Gruyter, Nueva York.

Batson, D. y L. Shaw, 1991, "Evidence for Altruism: Towards a Pluralism of Prosocial Motives", *Psychological Inquiry*, vol. 2, no. 2, pp. 107–122.

Brinck, I. y P. Gärdenfors, "Co-Operation and Communication in Apes and Humans", *Mind and Language*, vol. 18, pp. 484–501.
Darwin, C., 1989/1877, *The Descent of Man*, en *The Works of Charles Darwin*, ed. Barrett y Freeman, vol. 21, New York University Press, Nueva York.
Dawkins, R., 1989/1976, *The Selfish Gene*, 2a. ed., Oxford University Press, Oxford.
De Waal, F., 1996, *Good Natured. The Origins of Right and Wrong in Humans and Other Animals*, Harvard University Press, Cambridge, Mass.
Dennett, D., 1995, *Darwin's Dangerous Idea: Evolution and the Meanings of Life*, Simon and Schuster, Nueva York.
Fehr, E. y U. Fischbacher, 2003, "The Nature of Human Altruism", *Nature*, vol. 425, pp. 785–791.
Fehr, E. y S. Gächter, 2002, "Altruistic Punishment in Humans", *Nature*, vol. 415, pp. 137–140.
Flack, J. y F. de Waal, 2000, "Any Animal Whatever. Darwinian Building Blocks of Morality in Monkeys and Apes", en Katz 2000, pp. 1–29.
Gintis, H., S. Bowles, R. Boyd y E. Fehr, 2003, "Explaining Altruistic Behavior in Humans", *Evolution and Human Behavior*, vol. 24, pp. 153–172.
Heyes, C., 1993, "Anecdotes, Training, Trapping and Triangulating: Do Animals Attribute Mental States?", *Animal Behaviour*, vol. 46, pp. 177–188.
Hurley, S., 2003, "Animal Action in the Space of Reasons", *Mind and Language*, vol. 18, no. 3, pp. 231–256.
Huxley, T.H., 1995/1893, "Evolution and Ethics", en Thompson 1995, pp. 111–150.
Katz, L.D. (comp.), 2000, *Evolutionary Origins of Morality. Journal of Consciousness Studies*, vol. 7, no. 1.
Kummer, H., 2000, "Ways Beyond Appearances", en Katz 2000, pp. 48–52.
McGinn, C., 1979, "Evolution, Animals and the Basis of Morality", *Inquiry*, vol. 22, pp. 81–99.
Papineau, D., 2003, "Human Minds", en A. O'Hear (comp.), *Minds and Persons*, Cambridge University Press, Cambridge, 2003, pp. 159–184.
Rosas, A., 2004, "Mindreading, Deception and the Evolution of Kantian Moral Agents", *Journal for the Theory of Social Behaviour*, vol. 34, no. 2, pp. 127–139.
———, 2002, "Psychological and Evolutionary Evidence for Altruism", *Biology and Philosophy*, vol. 17, pp. 91–99.
Singer, P., 1981, *The Expanding Circle: Ethics and Sociobiology*, Farrar, Straus and Giroux, Nueva York.
Skyrms, B., 1996, *Evolution of the Social Contract*, Cambridge University Press, Cambridge.
Sober, E., 1993, *Philosophy of Biology*, Westview Press, Boulder.

Sober, E. y S.D. Wilson, 1998, *Unto Others. The Evolution and Psychology of Unselfish Behavior*, Harvard University Press, Cambridge, Mass.
Stevens, J.R. y M. Hauser, 2004, "Why Be Nice? Psychological Constraints on the Evolution of Cooperation", *Trends in Cognitive Sciences*, vol. 8, no. 2, pp. 60–65.
Thierry, B., 2000, "Building Elements of Morality Are Not Elements of Morality", en Katz 2000, pp. 60–62.
Thompson, P. (comp.), 1995, *Issues in Evolutionary Ethics*, SUNY Press, Albany.
Tomasello, M. y J. Call, 1997, *Primate Cognition*, Oxford University Press, Oxford.
Trivers, R., 1971, "The Evolution of Reciprocal Altruism", *The Quarterly Review of Biology*, vol. 46, pp. 35–57.
Williams, B., 1985, *Ethics and the Limits of Philosophy*, Harvard University Press, Cambridge, Mass.
Williams, G.C., 1995/1988, "Huxley's Evolution and Ethics in Sociobiological Perspective", en Thompson 1995, pp. 317–349.
———, 1989, "A Sociobiological Expansion of *Evolution and Ethics*", en J. Paradis y G.C. Wiliams (comps.), *Evolution and Ethics*, Princeton University Press, Princeton, 1989.
Wrangham, R. y D. Peterson, 1996, *Demonic Males. Apes and the Origins of Human Violence*, Houghton Mifflin, Nueva York.

Recibido el 14 de febrero de 2005; revisado el 10 de junio de 2005; aceptado el 7 de septiembre de 2005.