

---

## LA EVOLUCIÓN DE LA MENTE Y EL COMPORTAMIENTO MORAL

### The Evolution of Mind and Moral Behavior

PABLO QUINTANILLA<sup>1</sup>, Ph. D.

<sup>1</sup>Profesor Principal de Filosofía. Departamento de Humanidades.  
Pontificia Universidad Católica del Perú. pquinta@puccp.edu.pe

Presentado el 17 de septiembre de 2009, aceptado el 27 de octubre de 2009, correcciones 12 de mayo de 2010.

#### RESUMEN

Esta contribución discute algunas correlaciones entre la evolución de las capacidades cognitivas y afectivas humanas –específicamente metarrepresentación, simulación y simpatía– y la evolución del comportamiento moral. La tesis central es que en el caso de seres humanos, homínidos y algunos primates superiores, el altruismo moral y las capacidades metarrepresentacional y de simulación –que se construyen sobre habilidades simpáticas– son funciones que se potencian mutuamente. Esto parece sugerir que se trata de funciones que coevolucionaron. La evidencia empírica sugiere que podemos encontrar en primates no homínidos formas básicas de comportamiento altruista y capacidad metarrepresentacional, pero no hay evidencia de comportamiento altruista moral en especies o individuos humanos que carezcan de esta capacidad. En el caso del desarrollo de los niños, la capacidad de comportamiento moral es directamente proporcional al desarrollo de las capacidades metarrepresentacional y de simulación. El texto también propone una secuencia evolutiva que culmina en el altruismo moral y discute los rasgos que serían centrales al comportamiento moral humano.

**Palabras clave:** altruismo, evolución, mente, moral, metarrepresentación.

#### ABSTRACT

This contribution addresses some correlations between cognitive and affective human capacities –specifically metarepresentation, simulation and sympathy– and the evolution of moral behavior. The main claim is that in humans, hominids and some superior primates, moral altruism, metarepresentation and simulation –which are build on sympathetic abilities– are functions that reinforce each other. This might suggest that they are functions that coevolved. Empirical evidence suggests that we can find in non hominid primates basic forms of altruistic behavior and metarepresentation, but there is no evidence of moral altruistic behavior in species or human individuals that lack this capacity. In children's development, moral behavior is directly proportional to the development of metarepresentation and simulation. The paper also suggests an evolutionary sequence that culminates in moral altruism and discusses the traits that would be central to human moral behavior.

**Key words:** altruism, evolution, mind, morality, metarepresentation.

#### I. LA NATURALEZA DEL ALTRUISMO MORAL

Las discusiones en biología teórica y filosofía moral han producido abundante bibliografía en relación a la naturaleza, génesis y posibilidad del altruismo. Desde el punto de vista biológico, un

individuo es altruista si tiene un comportamiento que mejora las posibilidades reproductivas o de supervivencia de otros individuos, a expensas de las suyas propias. La existencia indiscutible de este fenómeno presenta, sin embargo, algunas preguntas. En primer lugar, todos los seres vivos estamos programados para luchar por nuestra supervivencia o, en todo caso, por la de individuos que comparten nuestros genes. Sin embargo, con frecuencia nos comportamos de manera altruista tanto con quienes comparten nuestros genes como con quienes no los comparten. ¿Por qué ocurriría esto? En segundo lugar, aparece aquí una paradoja: si el comportamiento altruista es un rasgo hereditario, y dado que los individuos altruistas se ponen a sí mismos en circunstancias menos ventajosas que los egoístas, sería de esperar que la selección natural vaya eliminando progresivamente a los altruistas hasta su total desaparición. Desde Darwin en adelante se han propuesto varias teorías para explicar estos problemas.

En este punto es importante notar que en la tradición filosófica los conceptos de “comportamiento altruista” y “comportamiento moral” no tienen el mismo significado ni son coextensivos. El primero alude al comportamiento que privilegia el beneficio ajeno por sobre el propio; el segundo alude al comportamiento caracterizado por una normatividad libremente elegida, lo cual implica algún grado de consciencia, libre albedrío y concepción del deber. Es claro que podría haber altruismo biológico sin moralidad, pero es debatible si podría darse moralidad sin altruismo, aunque yo me inclino porque no podría darse.

Sobre estos temas vienen discutiendo filósofos y biólogos los últimos años con resultados dispares. De un lado están aquellos que creen que es posible explicar tanto el comportamiento altruista como el comportamiento moral como una evolución del comportamiento cooperativo, es decir en términos naturalistas.<sup>1</sup> De otro lado están los escépticos ante la posibilidad de explicar la autonomía y la normatividad como una evolución de estados físicos prenormativos anteriores.<sup>2</sup>

Mi impresión, sin embargo, es que la discusión entre biólogos y filósofos con frecuencia se ve oscurecida porque se confunde altruismo y moralidad, por una parte, y porque los usos que hacen de la palabra altruismo no son equivalentes. El problema central que interesa a los biólogos es cómo pudo el altruismo biológico haberse originado y mantenido por selección natural. El problema que interesa a los filósofos, por su parte, suele ser si la teoría de la selección natural puede llegar a explicar algunos aspectos del comportamiento moral humano como, por ejemplo, las acciones altruistas entendidas como realizaciones autónomas, intencionales, deliberadas, conscientes y motivadas por creencias y deseos, y gobernadas por una normatividad elegida. Sin duda, ambas preguntas no son las mismas.

En este texto me propongo hacer dos cosas: 1. haré un análisis del concepto de altruismo para intentar aclarar sus diversos usos y 2. abordaré la pregunta acerca de cómo puede el estudio de la evolución de las funciones cognitivas ayudar a entender la evolución del altruismo y de la moralidad a partir de funciones naturales anteriores. Para lo primero distinguiré los diversos sentidos de altruismo así como las teorías que pretenden explicarlo, y sostendré que en el sentido filosófico moral del término, diferente del biológico, la atribución de comportamiento altruista moral a un individuo, es posible solo si se le atribuye también motivaciones conscientes y deliberadas. Para lo segundo, sostendré que el comportamiento moral en sentido estricto, que a mi parecer incluye comportamiento altruista biológico, solo es posible en un individuo que posee capacidad de simulación, una teoría de la mente (la llamada capacidad metarrepresentacional o mentalización) y sentimientos morales, así como deliberación acerca de las consecuencias de los actos.

---

<sup>1</sup> De este lado tienden a estar, por ejemplo, de Waal, Franz y Singer.

<sup>2</sup> De este lado están los filósofos de tendencia kantiana como Korsgaard (2006).

Es necesario hacer una precisión terminológica en este punto. Simulación es la capacidad de imaginar los estados mentales ajenos, en condiciones contrafácticas, como si fueran propios. Solo hay un matiz de diferencia con el concepto de metarrepresentación, entendido como la capacidad para representarse los estados mentales ajenos. El concepto de metarrepresentación está asociado a la posesión de una teoría de la mente y, por tanto, suele vincularse a la llamada teoría de la teoría (T-T) en el debate sobre la comprensión. Tener una teoría de la mente es poseer un conjunto de principios generalizadores, formulados en términos de regularidades nomológicas, que gobiernan las relaciones causales entre los estados mentales y las acciones de las personas. Cuando un niño ya posee una teoría de la mente, alrededor de los tres años, está en condiciones de explicar el comportamiento de otras personas encontrando relaciones causales, gobernados por regularidades nomológicas, a los estados mentales y acciones que reconoce en ellos. La teoría de la simulación (T-S), en cambio, afirma que las habilidades simulativas, en tanto la habilidad para imaginar contrafácticamente los estados mentales ajenos, son suficientes para la comprensión. En principio metarrepresentación y simulación no son incompatibles entre sí, aunque han dado lugar a un debate como si lo fueran (Dennett, 1978; Premack y Woodruff, 1978; Leslie, 1987; Frith *et al.*, 1991; Perner, 1991; Gordon, 1995a; Gordon, 1995b; Langdon *et al.*, 1997; Brunsteins, 2004; Brunsteins, 2008; Quintanilla, 2008). Me inclino porque la metarrepresentación presupone la simulación; además me parece que lo primero está caracterizado por funciones más cognitivas, mientras que lo segundo tiene bases prioritariamente afectivas. Sin embargo, no discutiré ese punto aquí. Para los fines de esta exposición asumiré que metarrepresentación y simulación se dan más o menos simultáneamente, aunque probablemente la simulación sea ligeramente anterior, y que tienen un origen evolutivo semejante. Lo más probable, aunque por el momento no hay pruebas conclusivas de ello, es que versiones menos evolucionadas de estas capacidades también estén presentes en primates no humanos. Pero lo que me interesa sugerir es que las capacidades metarrepresentacional y de simulación no solo potencian el desarrollo moral sino son condición de posibilidad de este.

Estas capacidades tienen una alta función de supervivencia porque permiten predecir el comportamiento de otros individuos, lo que es central para el comportamiento cooperativo en la defensa y la caza. Asimismo, la simulación está asociada a un simulador interno que le permite al agente imaginar escenarios alternativos realistas para poder actuar exitosamente. Esta capacidad también haría posible que el individuo delibere, realizando cálculos acerca de las mejores consecuencias de sus actos en términos del mayor beneficio individual y grupal. En términos morales, la atribución de estados mentales a otros individuos con los cuales uno se sentiría moralmente comprometido comenzaría con los individuos genética y socialmente más cercanos a uno y se ampliaría, como el círculo en expansión de Singer, 1981, y la torre de la moralidad de De Waal, 1997, hasta incluir otras o incluso todas las criaturas vivas.

## II. LAS EXPLICACIONES DEL ALTRUISMO

Son conocidas las diversas teorías que han sido propuestas para explicar la paradoja del altruismo. La teoría clásica procede de los capítulos cuarto y quinto de *The Descent of Man de Darwin* (Darwin, 1969), y suele recibir el nombre de selección de grupos (*group selection*). Según esta teoría son los grupos más cooperativos entre sí los que tienen mejores posibilidades de supervivencia respecto de grupos menos cooperativos, con lo cual la lucha por la supervivencia seleccionaría el rasgo altruista que a su vez se transmitiría hereditariamente. Es debatible si Darwin resolvió el problema de cómo llegan los individuos más cooperativos a sobrevivir frente a la competencia de individuos egoístas, antes que se conforme el grupo cooperativo. Para resolver ese problema surgieron nuevas propuestas. Hamilton, 1964, propuso la teoría de la selección de parentesco (*kin selection*),

---

según la cual la naturaleza ha seleccionado individuos altruistas porque esto favorece la supervivencia de las familias, es decir de los genes compartidos. Trivers, 1971, propuso la teoría del altruismo recíproco, para la cual es ventajoso para un individuo realizar actos altruistas pues estos generarán en los demás miembros del grupo actos igualmente altruistas respecto de ese individuo. Esta teoría se enriqueció con las investigaciones en teoría de juegos y, sobre todo, con los modelos basados en el comportamiento racional en el caso del dilema iterado de los prisioneros, según el cual si bien en casos aislados lo racional sería no cooperar, en casos iterados lo racional sí sería cooperar (Axelrod, 1994). A esto está asociado el modelo llamado del *tit for tat* (algo así como “toma y dame”), desarrollado por el psicólogo *Anatol Rapoport*, popularizado por Axelrod, y que consiste en cooperar en la primera movida y luego imitar la movida realizada por el oponente. De esta manera se coopera como reacción a la cooperación ajena o se actúa retaliatoriamente como reacción a la falta de cooperación ajena, pero se está dispuesto a cooperar posteriormente como consecuencia de actos cooperativos ajenos.

Posteriormente se produjo un renacimiento de la teoría de la selección de grupos gracias a Sober y Wilson, 1998, pero son muchos los autores que han aportado elementos a una concepción más comprensiva de la génesis del altruismo que incluye rasgos de más de una de estas teorías. Podría ser que los rasgos altruistas fueran seleccionados por razones tan diferentes como las que señalan las teorías de parentesco y altruismo recíproco, pero también porque el comportamiento altruista es un buen indicador de la calidad genética del individuo, porque es más eficiente compartir las grandes presas con rivales hambrientos antes que gastar energía defendiéndolas, porque los individuos egoístas son castigados por el grupo mientras los altruistas son premiados, etc. Rosas, 2004, ha denominado “selección psicológica”<sup>3</sup> al proceso por el que la naturaleza selecciona individuos altruistas a través de causas psicológicas, como por ejemplo el temor a la censura, el deseo de reconocimiento social, etc. (Darwin, 1969). En todo caso, no hay razón para suponer que estas teorías se excluyan entre sí y no puedan complementarse.

Si bien estas teorías pueden ser exitosas para explicar la aparición de altruismo biológico, se enfrentan con problemas frente al tipo de altruismo moral humano. Este ocurre característicamente cuando uno es altruista con un individuo que no pertenece a su grupo, con quien no se comparte carga genética, que no está en condiciones de actuar recíprocamente y en una situación en la que no hay ningún premio ni censura social en juego. Se podría discutir si este tipo de comportamiento altruista existe, aunque yo me inclino por pensar que sí, pero el punto es que si así fuera sería el único tipo de comportamiento verdaderamente altruista pues no tendría ninguna otra motivación excepto el bienestar ajeno como un fin en sí mismo.

Me parece claro que es posible encontrar este tipo de comportamiento en seres humanos. Se podría objetar que en los casos excepcionales en que éste se da, está motivado por la satisfacción narcisista de ayudar a otro o de evitar la culpa por no hacerlo. Sin embargo, eso solo probaría la existencia del altruismo moral, porque un individuo egoísta que actuara de manera altruista solo para obtener resultados que lo beneficiarían, no tendría ninguna satisfacción psicológica por beneficiar a alguien desinteresadamente, ni sentiría culpa por no hacerlo; en su caso sería todo lo contrario: se sentiría culpable respecto de sí mismo si actuara de manera desinteresada y no obtendría satisfacción psicológica de ello.

Aunque no hay todavía investigaciones exhaustivas al respecto, han sido reportados casos de comportamiento altruista moral en chimpancés frente a animales enfermos, agónicos o tullidos,

---

<sup>3</sup>Rosas remite a *El origen del hombre* de Darwin donde se encuentran explicaciones psicológicas de este tipo, aun cuando Darwin mismo no use la expresión “selección psicológica”. Cf. Darwin, *El origen del hombre*, op. cit., p. 136-137.

---

con los que no se comparte carga genética y que no pueden devolver el favor. Por otra parte, no es beneficioso para la especie preservar los genes de individuos enfermos o que simplemente no pueden perpetuar sus genes, con lo cual se mantiene la pregunta de por qué la naturaleza habría seleccionado a individuos que se comportan de manera desinteresadamente altruista, es decir, según lo que hemos venido llamado altruismo moral.<sup>4</sup> En el caso de ciertos animales, como por ejemplo las ratas, esto quizá sugiera que una vez que el rasgo altruista se ha instalado genéticamente actúa ciegamente, es decir, mecánicamente, independientemente de la función de supervivencia que pudiera tener para el individuo o el grupo. Pero mi punto es que si bien es cierto que el comportamiento cooperativo humano tiene una base genética similar a la de otros animales sociales y, por tanto, sus raíces pueden explicarse de la misma manera, hay una importante diferencia entre el comportamiento cooperativo animal y el comportamiento moral humano. Esta diferencia radica en que en el caso humano hay consciencia de que el acto altruista es desinteresado y normativo porque tiene como objetivo el bien de otro individuo que es considerado un fin en sí mismo. La posibilidad de llegar a desarrollar este tipo de consciencia parece requerir de capacidad de simulación. No me parece imposible que haya alguna forma primitiva de altruismo moral en animales no humanos, como ciertos primates superiores, pero esto es algo insuficientemente estudiado. Sí me parece claro que por complejos que sean los rasgos que constituyen el altruismo moral humano, y el comportamiento moral humano, tienen que ser el producto de la evolución de estados cooperativos menos complejos. Mi tesis, sin embargo, es que para que esta evolución se haya podido dar, es necesario que estén presentes facultades cognitivas muy complejas, especialmente la capacidad de simulación. Pero para aclarar estas ideas será necesario analizar los distintos sentidos de los términos involucrados.

### III. ALTRUISMO Y COMPORTAMIENTO MORAL

Hay cuatro niveles en la evolución del comportamiento moral que deseo describir. Cada uno de ellos está presente en el nivel posterior, es decir, el último nivel presupone a los tres anteriores y así en adelante.

En primer lugar tenemos el comportamiento cooperativo: este está caracterizado por ser un tipo de ayuda, colaboración o asistencia que se realiza para obtener un fin colectivo que beneficia al grupo. Hay comportamiento cooperativo en todos los animales sociales, incluidos los insectos. En segundo lugar tenemos el altruismo biológico: Aquí se trata de un tipo de colaboración que beneficia las posibilidades de supervivencia o reproducción de otro individuo a expensas de las de uno mismo. En tercer lugar está el altruismo moral: En este caso se trata de una colaboración libre, consciente, intencional y desinteresada que beneficia a otro individuo (con el que se puede o no compartir carga genética) a expensas de los intereses de uno mismo, donde el bienestar del otro es considerado un fin en sí mismo. Finalmente tenemos el comportamiento moral, que se caracteriza por ser un tipo de altruismo moral donde el agente considera que sus acciones están regidas por normas morales que él ha elegido. Se verá, entonces, que puede haber altruismo moral sin que haya consciencia del deber; esto ocurre, por ejemplo, cuando el agente actúa solo por compasión. Pero incluso en el caso del altruismo moral ya debe estar presente cierta capacidad de deliberación. Pero para que el agente actúe moralmente en sentido pleno, debe creer que se encuentra constreñido para actuar de esa manera, lo que le permite formular juicios morales.

---

<sup>4</sup> Un caso particularmente curioso es el siguiente experimento con ratas. En una jaula se encuentra un grupo de ratas que han aprendido a jalar una palanca para conseguir comida. En una jaula vecina hay una sola rata que recibe una descarga eléctrica cada vez que las otras ratas jalar la palanca de la comida. Después de algún tiempo, las ratas dejan de jalar la palanca aún si pasan hambre para evitar que la rata vecina sufra las descargas eléctricas.

---

En esta línea de pensamiento, lo que convierte a un acto en moralmente altruista o egoísta no es sus consecuencias respecto del bienestar de otros individuos, sino es el tipo de motivaciones que lo causaron, es decir, las creencias y deseos conscientes e intencionales del agente. En base a este supuesto es posible afirmar que solo criaturas a las que es posible atribuir creencias y deseos intencionales y conscientes son capaces de comportamiento moral. Es frecuente que estas motivaciones sean combinadas, es decir, que haya motivaciones altruistas integradas con egoístas. Sin embargo, basta con que haya alguna motivación desinteresada en la realización de una acción para que amerite considerarla como moralmente altruista y para que sea necesario preguntarse sobre su génesis y constitución. En todo caso, asumiendo que el comportamiento moralmente altruista se caracteriza porque sus motivaciones son desinteresadas, se está sosteniendo que quienes realizan estas acciones son criaturas deliberantes. La deliberación participa en dos momentos: en primer lugar para establecer que la acción moralmente correcta es aquella que beneficia a otros individuos a expensas de uno mismo, si es que el sacrificio que hay que hacer es inferior al beneficio que este proporcionará al otro individuo, con lo cual el bienestar del otro individuo se torna en algún sentido preferible al de uno y el bien del otro es visto como un fin en sí mismo. En segundo lugar, para establecer cuáles exactamente son las acciones que serían de beneficio de la otra persona. Así, entonces, la acción moralmente altruista incluye una deliberación consecuencialista.

Quizá haya quienes no encuentren razones para aceptar que el comportamiento moral es diferente del altruismo moral, y consideren que la noción de comportamiento moral que estoy empleando tiene un elemento de universalidad formal kantiana que no es necesario, ni conveniente, aceptar. La diferencia central que veo entre altruismo moral y comportamiento moral es que, *en principio*, lo primero podría darse en primates no homínidos (o en individuos humanos como niños o personas con problemas neurológicos) mientras que lo segundo no, porque este último implica capacidades cognitivas muy complejas. Sin embargo, considero que lo que tenemos es un continuo entre todos estos niveles, con lo que algunos rasgos de lo que llamamos comportamiento moral podrían estar presentes en especies o individuos de manera incipiente. Por otra parte, no estoy sosteniendo que haya algún sentido en que el comportamiento moral, así definido, sea superior o más valioso, en alguna escala de desarrollo moral, que el altruismo moral. Solo uso la categoría porque me parece que refleja lo que los filósofos morales comúnmente entienden por comportamiento moral, y me interesa mostrar su continuidad evolutiva con otras formas de comportamiento.

Ahora bien, intencionalidad, deliberación, consciencia de las motivaciones, libre albedrío y autonomía también son un asunto de grado y forman una continuidad. Es decir, en general podemos atribuir estas propiedades a las acciones realizadas por seres humanos adultos, pero en tanto vamos retrocediendo en la vida de la persona hasta su infancia (o en la evolución de nuestra especie desde otros homínidos), reconoceremos cada vez menos de estos rasgos en su comportamiento. No todo acto intencional es plenamente consciente (considérese los actos fallidos y los *lapsus* freudianos) y no todo acto consciente es plenamente intencional (a veces actuamos manipulados o bajo presión).<sup>5</sup> Pero consciencia e intencionalidad van de la mano, incluso si en proporciones diferentes. De otro lado, consciencia e intencionalidad también forman un continuo evolutivo, de suerte que podemos encontrar una graduación que va desde el *homo sapiens*, pasando por los primates no humanos, hacia los mamíferos superiores e inferiores. Naturalmente una graduación semejante, no idéntica, la encontramos en el desarrollo del niño. La capacidad de deliberar acerca de las consecuencias de nuestras acciones también es un asunto de grado, con lo cual es

---

<sup>5</sup> Si bien todo acto es por definición intencional, la intención puede ser consciente o inconsciente, en un sentido psicoanalítico. Uno, por ejemplo, podría realizar una acción motivado por causas que no conoce, pero seguiría siendo una acción porque habría habido intencionalidad, aunque fuera inconsciente.

---

razonable suponer que lo es el comportamiento moral en general, teniendo como criterio básico de comparación el comportamiento moral humano. En su versión humana desarrollada, desde por lo menos el altruismo moral, estas capacidades parecen inseparables de la capacidad de simulación; en tanto esta capacidad está menos presente, también lo está el altruismo y el comportamiento moral. Así pues, ahora volveremos nuestros ojos a la capacidad de simulación.

#### IV. SIMULACIÓN Y METARREPRESENTACIÓN

La teoría más aceptada acerca de la evolución cognitiva es el modelo de autocatálisis. Según este, cuando los homínidos de hace aproximadamente dos millones de años se volvieron bípedos, como consecuencia de cambios en el ambiente, sus extremidades anteriores quedaron liberadas, con lo cual pudieron ser usadas para la confección y uso de artefactos. El uso de utensilios potenció la inteligencia y aumentó el volumen de masa encefálica, de aproximadamente 450 cm<sup>3</sup> hace tres millones de años (comparable a los cerebros actuales de chimpancés y gorilas), hasta aproximadamente 1.500 cm<sup>3</sup> en el *Homo sapiens sapiens* actual. También se desarrolló y perfeccionó el comportamiento cooperativo para la caza y la protección mutua, lo que desarrolló más aún la inteligencia.

Las habilidades involucradas en el comportamiento cooperativo para la caza y la protección mutua tienen que haber evolucionado de la mano con mecanismos de reconocimiento de las necesidades, emociones, creencias y objetivos ajenos. Estos mecanismos requieren de la aparición de capacidades simpáticas, empáticas y metarrepresentacionales. Por razones metodológicas seguiré una tendencia actual a distinguir entre simpatía y empatía (Wispé, 1987; Wispé, 1991). Lo primero alude a la capacidad de percibir y compartir, o acompañar emocionalmente, los estados mentales ajenos. Este concepto, popularizado filosóficamente por Hume, se encuentra muy cercano al de compasión, pues tiene un contenido básicamente afectivo y se puede reconocer ya en niños pequeños, así como en primates no humanos cognitivamente poco desarrollados.

La empatía, por su parte, es una habilidad más cognitiva que afectiva, y alude especialmente a la capacidad de reconocer las creencias ajenas y sus necesidades y sentimientos, incluso si uno no los comparte, si a uno no lo movilizan psicológicamente o si uno emplea ese conocimiento precisamente para manipular el comportamiento ajeno. Hay debate respecto de si se puede separar simpatía y empatía, y es conveniente aclarar que estas suelen ser habilidades que se dan juntas, pero es conveniente distinguirlas conceptualmente para iluminar los distintos rasgos involucrados. Por otra parte, al parecer sí pueden darse separadamente. Es posible encontrar personas con gran capacidad empática pero mínima o ninguna capacidad simpática, como es el caso de los psicópatas. También está reconocida la existencia de individuos (sobre todo bebés muy pequeños e incluso animales) que son capaces de percibir el sufrimiento ajeno y sentirse afectados por él, sin tener un desarrollo cognitivo suficientemente complejo como para entender, o siquiera imaginar, la situación por la que la otra persona o criatura está pasando, es decir, sin estar en condiciones de hacer atribuciones de estados mentales en segundo grado o, lo que es lo mismo, sin tener capacidad metarrepresentacional. Entonces, podemos distinguir entre empatía y simpatía como capacidades diferentes que ciertamente se complementan y se potencian mutuamente. Todo parece indicar, sin embargo, que la simpatía es evolutivamente anterior a la empatía.

Formas rudimentarias de la simpatía pueden ser encontradas en mamíferos superiores y en bebés muy pequeños, con incluso solo horas de vida, en lo que se conoce como “contagio emocional” (Hartfield *et al.*, 1993).<sup>6</sup> Esto se produce cuando un individuo reconoce estados mentales

---

<sup>6</sup> Según estos autores el contagio emocional comienza como una tendencia automática a imitar y sincronizar con otra persona posturas y gestos, convergiendo posteriormente en el plano emocional.

en otro, especialmente aquellos que involucran sufrimiento pero también placer o dicha, reproduciéndolos en sí mismo. Esta reproducción suele ser inconsciente y precognitiva. Un ejemplo de esto es cuando tenemos un grupo de recién nacidos y uno de ellos se pone a llorar; lo más probable es que todos los demás se contagien de esto y rompan también a llorar, incluso si no tienen ninguna incomodidad. El contagio emocional tiene un alto valor de supervivencia porque permite reaccionar apropiadamente ante situaciones compartidas como, por ejemplo, con temor y huida ante una amenaza común. Probablemente el sustrato más básico del contagio emocional sea el motor *mimicri*, o imitación mecánica del comportamiento ajeno. Esto es lo que ocurre, por ejemplo, con una manada de alpacas que mecánicamente dejan lo que están haciendo y se ponen a mirar en cierta dirección, una vez que una de las alpacas ha actuado de esa manera. *Motor mimicri* y contagio emocional son probablemente las raíces evolutivas de lo que después serán la simpatía, la empatía y la simulación, que originalmente tendrían como objetivo permitir que los individuos se apoyen mutuamente en la satisfacción de necesidades y se protejan de peligros comunes.

En lo que sigue voy a imaginar una secuencia evolutiva que conduciría al altruismo y comportamiento moral, donde cada paso anterior es condición de posibilidad del posterior. Los conceptos entre paréntesis son explicaciones o se dan de manera asociada a los que vienen inmediatamente antes de ellos. Después de exponer el esquema intentaré justificarlo.

Comportamiento cooperativo, altruismo biológico, *motor mimicri*, contagio emocional → simpatía (compasión) → empatía (reconocimiento de la propia identidad en un espejo, consciencia de sí mismo, identificación con otro) → metarrepresentación/simulación (teoría de la mente, capacidad de mentalización, capacidad de atribuir estados mentales en segundo o más grados), capacidad de postergación → altruismo moral → deliberación, reconocimiento de intencionalidad moral, interpretación y juicio moral, autonomía → comportamiento moral.

En lo que sigue voy a sugerir algunas correlaciones entre la evolución cognitiva y moral del niño, así como entre la evolución cognitiva y moral de la especie. No se trata, por supuesto, de suponer que la ontogénesis recapitula la filogénesis, sino de iluminar posibles líneas similares de evolución.<sup>7</sup>

La evolución empática del niño toma más o menos la siguiente forma: Antes de cumplir el año los infantes ya distinguen entre criaturas vivas y objetos inertes. A partir del año los niños consuelan a los demás. El que esto sea un comportamiento tan precoz sugiere su carácter innato y, por tanto, adaptativo y no cultural. Esto sugiere que en el desarrollo moral primero se forman las tendencias afectivas y después aparecen las racionalizaciones del comportamiento generado por ellas, todo lo cual no haría sino apoyar la célebre frase de Hume de que la razón está y debe estar al servicio de los afectos.

Aproximadamente a los catorce meses los niños pueden involucrarse en situaciones de atención compartida (*joint attention*) con otras personas acerca de objetos comunes, hacen contacto visual y pueden seguir la dirección de las miradas ajenas. Así, por ejemplo, si el bebe ve que la madre mira hacia algún objeto distante con interés, el niño sigue la mirada de la madre y reacciona ante las reacciones de ella. Estos son los rudimentos de lo que más adelante será la capacidad de triangular, en que el niño desarrolla los conceptos de yo, otros y mundo compartido, como conceptos inseparables.

El contagio emocional progresivamente se va convirtiendo en capacidad simpática y empática al asociarse con la capacidad de identificación y la consciencia de sí mismo, lo que suele

---

<sup>7</sup> Dos variables están correlacionadas cuando los valores de una de ellas varía sistemáticamente en función a los valores de la otra. Así, hay una correlación entre las capacidades cognitivas y morales de un niño (o de una especie) cuando se observa que los valores de las capacidades cognitivas y del comportamiento moral varían sistemáticamente según las variaciones de la otra.



considerarse que ocurre cuando el niño puede reconocerse en un espejo, entre los 18 y los 24 meses. El retraso de la capacidad de reconocerse en un espejo está relacionada con el retraso en la capacidad de atribuir estados mentales a otros individuos, lo que suele ocurrir en niños con retardo mental, autismo y daño en el lóbulo frontal (Hauser, 2006).

Antes de los dos años aparece la capacidad conocida como *pretend play* o juego simbólico. Por ejemplo, el niño jugará a que el sillón de la sala es un caballo o a que su muñeca es un bebé, pero claramente reconocerá que esto es solamente un juego y distinguirá entre la realidad y su fantasía. Al jugar, el niño imagina que la realidad es diferente de como él sabe que es. Aquí se establecen las bases del principio de realidad y de lo que posteriormente serán las capacidades de simulación y metarrepresentación.

Aproximadamente a los dos años de edad el niño desarrolla la capacidad de representarse el mundo, con lo cual puede diferenciar la realidad de la fantasía y, así, adquiere los rudimentos de lo que después serán los conceptos de verdadero y falso. Alrededor de los tres años, el niño desarrolla la capacidad de representarse las representaciones ajenas (la capacidad metarrepresentacional), con lo cual puede atribuir a otros agentes estados mentales diferenciándolos de los suyos propios (“Pedro cree que p”) y está en condiciones de aprobar las pruebas de la falsa creencia (Fonagy *et al.*, 2002).<sup>8</sup> Recién a los cinco o seis años el niño está en capacidad de atribuir a otras personas estados mentales de segundo orden (“Pedro cree que María cree que p”). De esta manera, el niño va aprendiendo a compartir el punto de vista ajeno sin perder el propio.

Las investigaciones de Fonagy *et al.*, 2002, indican que si bien todos los niños desarrollan de manera espontánea una teoría de la mente, es decir, la capacidad de atribuir estados mentales a los demás para explicar su comportamiento, las capacidades de simulación varían notablemente según la calidad del apego que los niños tuvieron con sus cuidadores tempranos (Bowlby, 1973; Bowlby, 1978). Estos mismos autores denominan “mentalización” o “función reflexiva” a la capacidad de tener estados mentales y de atribuirlos a otros individuos. Según Fonagy *et al.*, 2002, un apego deficiente generará una mala función reflexiva, lo que en situaciones extremas podría conducir al individuo a ver a las demás personas de manera deshumanizada, como miembros de una jerarquía o grupo, y no como agentes intencionales dotados individualmente de estados mentales.

Entre los dos y los cuatro años el niño puede comenzar a imaginar de manera básica los sentimientos y estados mentales ajenos. Mucho antes de desarrollar capacidades metarrepresentacionales, los niños pequeños ya tienen vínculos afectivos con otras personas, lo que sugiere que la metarrepresentación se desarrolla desde habilidades preconceptuales no-conscientes que se van haciendo conscientes y desarrollando paulatinamente, a partir de factores neurológicos y de condiciones del entorno como el apego. Todas estas habilidades son necesarias para la supervivencia. Cuando Darwin estudió el comportamiento cooperativo en *The Descent of Man* afirmó que es un tipo de simpatía, aludiendo al término usado por Hume y Adam Smith, a

---

<sup>8</sup> Estas pruebas funcionan de la siguiente manera. En un teatro de marionetas se tiene a los personajes A y B. A juega con una canica y la deja en una canasta de mimbre, para luego salir del escenario. B saca la canica de la canasta y la deja en una caja de cartón, después también sale del escenario. A entra nuevamente al escenario sin saber que B ha cambiado de lugar la canica. El experimentador pregunta a dos grupos de niños, un grupo de alrededor de 3 años y medio y otro grupo de cerca de 5 años, dónde buscará A la canica. Los niños más jóvenes contestarán que A la buscará en la caja de cartón (donde realmente está), mientras que los niños mayores dirán que A la buscará en la canasta de mimbre (donde A la puso). El experimento muestra que el grupo más joven carece aún de capacidad metarrepresentacional al no poder atribuir a A una creencia que ellos consideran falsa, pues eso supondría alejarse de sus propias creencias y ver el mundo desde otra perspectiva. Los niños mayores de cinco años, en cambio, no tienen ningún inconveniente en atribuir a A la creencia falsa y, por tanto, en representarse las representaciones de A. Este experimento prueba que ellos pueden representarse el mundo desde el punto de vista de A, sin perder su propia perspectiva de las cosas.

---

quienes leyó y apreció, añadiendo que esto permitiría en el caso de animales más evolucionados la aparición de una “consciencia o sentido de lo moral”. Lo importante es que las habilidades metarrepresentacionales se desarrollan cuando el niño tiene que atribuir a los demás estados mentales para poder comprenderlos y explicar coherentemente su comportamiento. De esta manera se habitúa a simular ser el otro en condiciones contrafácticas, imaginando estados mentales que no tiene pero que podría haber tenido. Así, el niño se acostumbra a comportarse con los demás según los estados mentales que les va atribuyendo, que son los que el niño cree que tendría si estuviera en la situación que él cree que están los otros. Algunos estudios (Hauser, 2006) muestran una correlación entre la capacidad metarrepresentacional y el juicio moral. A los cinco años los niños juzgan los actos moralmente, no según sus consecuencias sino dependiendo de las intenciones que los niños creen que tuvieron los agentes que las realizaron. Así, los niños juzgan de maneras diferentes actos que tuvieron exactamente las mismas consecuencias, si es que los niños atribuyen intenciones distintas a sus agentes. A esa edad los niños no solo atribuyen intenciones a las personas sino, además, les atribuyen intenciones que juzgan como buenas o como malas. Es claro que para hacer ese tipo de juicio moral los niños necesitan haber imaginado los estados mentales del agente (su buena o mala intención), lo que parece probar que el juicio moral presupone cierta capacidad metarrepresentacional.

Las capacidades metarrepresentacionales y de simulación tienen una obvia función adaptativa, pues permiten realizar de manera exitosa actividades cooperativas, como la caza y la defensa, elaborando tácticas comunes. También permiten engañar y manipular, pero además hacen posible la transmisión de información y la construcción de un orden social que tenga complejas relaciones de jerarquía y poder, así como de solidaridad, reciprocidad y altruismo. Es posible afirmar que hay una relación directamente proporcional entre la presencia de capacidad metarrepresentacional en una especie, y la complejidad de su estructura social y de los vínculos intersubjetivos que se dan en ella. Veremos más adelante que hay razones para suponer que existen capacidades metarrepresentacionales en otros primates superiores además de los humanos. En todo caso, la correlación entre complejidad social y capacidad metarrepresentacional sugiere que las habilidades cognitivas y afectivas humanas, que son las que participan en las capacidades metarrepresentacional y de simulación, han evolucionado modeladas según su función social (Byrne y Whiten, 1988; Cosmides, 1989). Esto es lo que se suele llamar la hipótesis de la inteligencia social o del cerebro social. Según Dunbar, las investigaciones neurológicas del neocórtex sugieren la hipótesis de que ya en los primates existiría un “cerebro social” (Dunbar, 1998; Obiols y Pousa, 2005).

El punto es que las capacidades metarrepresentacional y de simulación hacen posible la existencia de complejas relaciones sociales, entre estas está el reconocimiento del sufrimiento ajeno y una tendencia natural a aliviarlo, especialmente si se trata de un miembro del grupo más cercano. Este sería el origen del comportamiento altruista moral. Las investigaciones sugieren que el reconocimiento del sufrimiento ajeno produce un sufrimiento propio, lo que activa un comportamiento de cooperación. Existe este tipo de comportamiento en los seres humanos, aunque hay reportes documentados de comportamiento compasivo en primates no humanos y en mamíferos como ballenas, delfines, elefantes, lobos y perros (De Waal, 1997).

La evidencia en psicología experimental sugiere que el aumento de las capacidades simpática y empática potencia el altruismo moral. Asimismo, sugiere que los niños expuestos a cuidadores empáticos y previsibles, es decir racionales y cuyo comportamiento es inteligible, desarrollan más empatía y capacidad de introspección. Esto haría posible que esos niños se conviertan en adultos más altruistas (Fonagy *et al.*, 2002).

Las evidencias sugieren que hay una correlación entre capacidad metarrepresentacional, capacidad de postergación y altruismo (Thompson *et al.*, 1997; Moore y Macgillivray, 2004). Esto

---

es fundamental porque el comportamiento moral normalmente requiere que uno postergue, o eventualmente anule, sus deseos inmediatos en favor de un beneficio o recompensa posterior, que puede ser un beneficio propio, el equilibrio social o un beneficio para otros. Además, tanto la capacidad de postergación como el altruismo requieren de la habilidad para imaginar estados mentales futuros. En el caso de la capacidad de postergación, imaginamos los estados mentales que tendríamos si postergáramos cierta situación presente a cambio de un beneficio futuro; en el caso del altruismo, imaginamos los estados mentales futuros de otra persona si actuáramos de cierta manera para con él. Los niños que responden bien a las pruebas de la falsa creencia tienen una mayor tendencia a compartir con otros niños a cambio de una gratificación futura (Hauser, 2006).

Esto parece requerir de un simulador interno que permita al individuo imaginarse a sí mismo en un momento futuro, obteniendo o realizando para otros un beneficio que no se obtendría de no aceptar un perjuicio presente, o imaginando a otros gozando de un beneficio futuro que solo se daría si el agente hiciera un sacrificio presente. Un simulador interno es una facultad que nos permite imaginarnos a nosotros mismos en distintos escenarios posibles, lo que nos ayuda a decidirnos por la opción que presente mejores consecuencias. El simulador no solo permite que uno se imagine circunstancias del entorno, sino también estados mentales propios y ajenos. Un simulador interno de este tipo es altamente adaptativo porque permite probar imaginariamente distintos cursos de acción antes de realizarlos y correr el riesgo de fracasar. Ahora bien, lo esencial en el comportamiento moral no es solo ser capaz de comportamiento altruista, sino también hacerlo de manera voluntaria y como producto de algún tipo de deliberación moral, lo que incluye el reconocer intenciones (en uno mismo y en los demás) y valorarlas como buenas o malas. Esta capacidad parece presuponer todas las capacidades mencionadas que los niños van adquiriendo progresivamente.

#### V. ALTRUISMO, SIMULACIÓN Y METARREPRESENTACIÓN EN PRIMATES NO HUMANOS

Aunque es un hecho que los primates no humanos son capaces de comportamiento altruista biológico, no es posible atribuirles comportamiento moral en el sentido que venimos discutiendo, y es discutible si podemos atribuirles altruismo moral, pues eso implicaría atribuirles algún grado de capacidad metarrepresentacional y de simulación. Sin embargo, sí podemos encontrar en algunas especies las bases evolutivas de lo que en los humanos se convirtió en el altruismo moral.

Los experimentos clásicos realizados por Premack y Woodruff, 1978, y publicados en su artículo de 1978 en que acuñan la expresión “teoría de la mente”, exhiben ante un grupo de simios un video con personajes. La evidencia indica que los simios llegan a reconocer las intenciones de los individuos que aparecen en el video.

Por otra parte, chimpancés y gorilas pueden reconocer su propia imagen en un espejo, lo que es indicador de consciencia de sí mismo (Povinelli, 1995).<sup>9</sup> Chimpancés, gorilas y bonobos tienen la habilidad de predecir, engañar y manipular el comportamiento de otros individuos, así como de crear alianzas en contra de otros. Esto fue denominado por Humphrey “inteligencia maquiavélica” (Humphrey, 1976). Los chimpancés, además, tienen la habilidad de contraengañar a otros. El engaño es hacer creer a otro individuo algo que uno no cree. El contraengaño es hacerle creer al otro que uno cree algo que no cree. Esta habilidad requiere, por lo menos, de capacidad metarrepresentacional en segundo grado. Considérese la siguiente situación comentada por Byrne y Whiten, 1988 y Obiols y Pousa, 2005:

---

<sup>9</sup> Esto se prueba en condiciones experimentales poniéndole al animal un punto de color en la frente y mostrándole un espejo. El animal se tocará la frente reconociendo que la imagen reflejada es la de él mismo. Algunos animales también hacen gestos y piruetas e incluso se adornan la cabeza, coquetamente, con ramas y hojas mientras se miran al espejo.

El chimpancé A puede acceder a un cajón con comida.  
El chimpancé B no puede hacerlo, pero vigila a A.  
A simula no tener comida ni acceso a la comida.  
B simula irse, pero se esconde tras un árbol, mientras vigila a A.  
A saca la comida.  
B retorna rápidamente y arrebató a A la comida.

En este experimento A atribuye creencias a B, pero también atribuye a B creencias acerca de A. B hace lo mismo con A. Finalmente, B gana en el juego del engaño.

Otro experimento realizado por Povinelli (Povinelli, 1987; Povinelli, 1990; Povinelli, 1992; Povinelli, 1994) y De Waal, 1997, en el laboratorio de Sarah Boysen tenía las siguientes características: Dos personas diferentes señalaron a un grupo de simios dónde se encontraba escondido un paquete de comida. Los simios observaron que una de estas personas había visto dónde se escondió la comida; también vieron que la otra persona tenía la cabeza cubierta con una bolsa de papel cuando la comida fue escondida. Los simios confiaron en las indicaciones de la primera persona pero no en las de la segunda persona para ubicar la comida. Esto sugeriría que los chimpancés pueden tener éxito en la prueba de la falsa creencia. No hay acuerdo pleno en este punto, sin embargo, porque los experimentos se han realizado con simios acostumbrados al trato humano y no en estado silvestre. Premack mismo, diez años después de su artículo clásico publicado con Woodruff, concluyó que no hay pruebas suficientes para atribuir a los chimpancés una teoría de la mente (Premack y Woodruff, 1978). Sin embargo, los experimentos realizados por Kuroshima *et al.*, 2003, de la Universidad de Kyoto sugieren que algunos monos capuchinos tendrían éxito en la prueba de la falsa creencia. De Waal, 2006, piensa que los experimentos donde los chimpancés no parecen superar esta prueba estuvieron mal dirigidos. Según él, esos experimentos no tenían el elemento afectivo necesario para movilizar a los simios, por lo cual el problema con los experimentos es que estaban diseñados de una forma demasiado “cognitivamente humana” (Hammock y Young, 2005).<sup>10</sup> En todo caso, todo lo que se puede decir es que no tenemos razones categóricas para afirmar que los chimpancés tienen capacidad metarrepresentacional semejante a la humana. Más aún, es razonable suponer que la capacidad metarrepresentacional humana, plenamente desarrollada, requiere de lenguaje, habiendo estas dos capacidades evolucionado interactuando mutuamente. De ser así, la capacidad metarrepresentacional humana plenamente desarrollada tendría alrededor de 50.000 años, aunque tendría que haber evolucionado de capacidades cognitivas más básicas y mucho más antiguas.

Lo que sí podemos afirmar es que hay suficientes indicios de que los chimpancés tienen algún tipo de capacidad metarrepresentacional que les permite atribuir estados mentales a otros individuos, predecir su comportamiento, diseñar estrategias de engaño y monitorear el comportamiento ajeno.<sup>11</sup> Los chimpancés pueden atribuir estados mentales a otros individuos, pero se confunden cuando tienen que diferenciar entre individuos para atribuir distintos estados mentales. Esto sugeriría que poseen una versión primitiva de una teoría de la mente. En la medida que nos vamos alejando de los patrones evolutivos humanos, estas capacidades se hacen más tenues. Por ejemplo,

---

<sup>10</sup> De Waal también piensa que los estudios con bonobos podrían ser cruciales, dado que los bonobos son los primates más empáticos y que los estudios de ADN entre bonobos y humanos sugieren que compartimos con estos animales, pero no con los chimpancés, un microsatélite relacionado con la socialización. Cf. Hammock y Young (2005).

<sup>11</sup> Como se sabe, los primates se dividen en tres grandes grupos: monos del nuevo mundo (capuchinos, muriquis, monos ardilla), monos del viejo mundo (babuinos y macacos), y homínidos y simios, (orangutanes, gorilas, bonobos y chimpancés; siendo los últimos los más cercanos evolutivamente a los humanos). La separación evolutiva entre los chimpancés y los homínidos se dio hace aproximadamente cinco o seis millones de años, con lo cual la formas más básicas de capacidad metarrepresentacional deben tener algo más que ese lapso de tiempo.

---

los gorilas también se reconocen en espejos y pueden monitorizar la mirada de otros individuos, pero no hay evidencia de que atribuyan estados intencionales ni que tengan éxito en la prueba de la falsa creencia. Algunos monos son capaces de distraer a sus rivales comportándose como si, y por tanto haciéndoles creer que, hay un depredador acercándose, pero no se reconocen en espejos ni superan pruebas de la falsa creencia. Byrne y Whiten (Byrne y Whiten, 1988; Byrne y Whiten, 1991) denominan a este comportamiento en monos “engaño táctico”, y consideran que es un comportamiento rutinario y mecánico que no presupone capacidad metarrepresentacional. De ser así, ello indicaría que no todo comportamiento social complejo requiere de capacidad metarrepresentacional y que las formas más primitivas de esa capacidad se encuentran en comportamientos mecánicos más básicos.

Por otra parte, así como hay pruebas de capacidad metarrepresentacional en chimpancés pero no en monos, hay indicios de comportamiento compasivo y empático en los primeros pero no en los segundos. Toda esta evidencia interconectada sugeriría que la compasión está presente donde hay capacidad de simulación, y que estas son las condiciones básicas para el posterior desarrollo de juicio y comportamiento moral. Además, por lo menos en el caso de los niños, hay una correlación entre la capacidad de reconocerse a uno mismo en un espejo y la tendencia a compadecerse y consolar a quienes sufren. Esto sugeriría que el consuelo y la compasión requieren de consciencia de sí mismo y capacidad de imaginarse padeciendo lo que otro padece (Johnson, 1982; Bischof-Köler, 1988).

Como se ha visto, si bien empatía y simpatía están fuertemente integrados, metodológicamente podemos distinguirlas porque la primera tiene más carga cognitiva y la segunda más carga afectiva. Coincido con De Waal (De Waal, 2003; De Waal, 2006)<sup>12</sup>, Baron-Cohen (Baron-Cohen, 2003; Baron-Cohen, 2004) con otros investigadores en que a la base de la empatía y la capacidad metarrepresentacional se encuentra la simpatía, es decir, los lazos afectivos entre individuos de un mismo grupo. La simpatía, a su vez, reposa sobre mecanismos más básicos aun, como el contagio emocional y el motor mimicri. De hecho, cuando uno observa una fotografía donde se ve a una persona en extremo sufrimiento, por ejemplo con un rictus de dolor, la cara del observador tiende a imitar la de la imagen, y el observador reproduce en sí mismo la experiencia que supone tuvo el fotografiado. Así, por ejemplo, se genera un sentimiento inmediato de malestar, incomodidad, desagrado o, incluso, sufrimiento físico. Estas reacciones musculares son mecánicas y automáticas e, incluso, pueden ser inconscientes e invisibles al ojo no experimentado (Dimberg, 2000). El punto es que reconocer un estado mental en otro individuo implica, en algún sentido, compartirlo. No es necesario padecerlo plenamente pero sí imaginar cómo sería tenerlo y esto es, de alguna manera, experimentar algo de él. Esto sugiere que no puede haber empatía sin algún elemento simpático, aunque hay formas de empatía que llegan a desconectarse de lo afectivo y tienen una mayor carga cognitiva. Todo esto sugeriría una secuencia evolutiva donde los primates menos evolucionados poseen contagio emocional, capacidad simpática y alguna forma básica de capacidad empática y de simulación, más no llegan a tener capacidad metarrepresentacional. Sin embargo los primates más cercanos a los humanos, los chimpancés, sí llegan a tener formas rudimentarias de capacidad metarrepresentacional. En el caso de los homínidos, la capacidad metarrepresentacional se habría potenciado con el desarrollo del cerebro y con la complejidad de las relaciones sociales, dando lugar a formas de altruismo que ya no son solo biológicas sino también morales. Finalmente, con

---

<sup>12</sup> De Waal ha propuesto lo que él llama “el modelo de la muñeca rusa”. Según este modelo, la capacidad más básica y específica es el contagio emocional. Conteniendo a esta, como un desarrollo posterior, está la empatía cognitiva. Conteniendo a esta última, y por tanto también a la primera, está la plena capacidad de atribuir estados mentales a los otros.

la aparición del lenguaje, la capacidad metarrepresentacional humana se habría potenciado (y habría potenciado al lenguaje) a tal punto que habría conformado las bases para la aparición del comportamiento moral en su estado actual. Singer, 1981, y De Waal, 2006, coinciden en que el juicio moral, es decir la aprobación y desaprobación moral, solo está plenamente presente cuando hacemos juicios que tienen la pretensión de ser imparciales y generales, esto es, cuando intentamos salir de nuestra propia perspectiva para ver el comportamiento y las intenciones, nuestras o ajenas, desde un punto de vista objetivo, como si fuera desde un punto de vista privilegiado. Solo en esas circunstancias podríamos hablar, de manera auténtica, de comportamiento moral, y es claro que este tipo de comportamiento solo se da en seres humanos.

La tesis de este artículo, sin embargo, es que el comportamiento moral humano es el producto de la evolución de capacidades anteriores, muchas de las cuales pueden ser encontradas en sus formas rudimentarias en otros primates o mamíferos sociales. De ser así, estaríamos en camino de encontrar una explicación naturalista del comportamiento moral humano. Asimismo, la tesis sostiene que el altruismo moral (y por tanto también el comportamiento moral) involucra sentimientos de compasión, los que serían posibles gracias a las capacidades de simulación y metarrepresentación. La evolución nos ha provisto de estas capacidades que tienen un alto valor de supervivencia, lo cual permitiría la aparición de sentimientos morales. La presencia de estas capacidades y sentimientos hace posible el comportamiento altruista moral en individuos que no tienen la posibilidad de comportarse recíprocamente, o en situaciones que no representan ningún tipo de beneficio para el agente. El altruismo moral no parece tener necesariamente una función de supervivencia para el agente o para el grupo, pero es consecuencia de las capacidades anteriores. Se trataría de rasgos que, en el caso de los humanos, el desarrollo cerebral, la complejidad social y la cultura potenciarían. Estas habilidades se convertirían en plenas habilidades morales cuando el individuo está en condiciones de elaborar juicios morales con pretensiones de objetividad y de reconocer su comportamiento como constreñido por normas que él reconoce como valiosas y que acepta voluntariamente.

## BIBLIOGRAFÍA

- AXELROD R. The evolution of cooperation. En: Singer P, editor. *Ethics*. Oxford: Oxford University Press; 1994.
- BARON-COHEN S. *The Essential Difference*. Nueva York: Basic Books; 2003.
- BARON-COHEN S. Sex differences in social development: lessons from autism. En: Leavitt L.A., Hall D.M.B. *Social and Moral Development: Emerging Evidence on the Toddler Years*, Johnson and Johnson Pediatric Institute; 2004. p. 125-141.
- BISCHOF-KÖLER D. Über den Zusammenhang von Empathie und der Fähigkeit sich im Spiegel zu erkennen. *Schweiz Z Psychol*; 1988;47:147-159.
- BOWLBY J. *Attachment and Loss*. Londres: Hogart Press; 1973.
- BOWLBY J. The role of attachment in personality development. En: *A secure base: clinical applications of attachment theory*. Londres: Routledge; 1989.
- BRUNSTEINS P. Estrategias de atribución mental: algunas distinciones y combinaciones. En: Rabossi E, editor. *La mente y sus problemas. Temas actuales de filosofía de la psicología*. Buenos Aires: Catálogos; 2004.
- BRUNSTEINS P. Algunas reflexiones acerca de la simulación mental y la perspectiva de la tercera persona. *Areté*. 2008;20(1).
- BYRNE RW, WHITEN A, editores. *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes and humans*. Oxford: Clarendon Press; 1988.

- 
- BYRNE RW, WHITEN A. Tactical deception in primates. *Behav Brain Sci.* 1988;11:233-273.
- BYRNE RW, WHITEN A. The emergence of metarepresentation in human ontogeny and primate phylogeny. En: WHITEN A, editor. *Natural theories of mind.* Oxford: Basil Blackwell; 1991.p. 267-281.
- COSMIDES L. The logic of social exchange: has natural selection shaped how humans reason? *Studies with the Wason selection task.* *Cognition.* 1989;31:187-276.
- DARWIN C. *El origen del hombre y la selección en relación al sexo.* México: Editorial Diana; 1969. p. 136-137.
- DIMBERG U, THUNBERG M, ELHEMED K. Unconscious facial reactions to emotional facial expressions. *Psycholog Sci.* 2000;11:86-89.
- DENNETT D. Beliefs about beliefs. *Behav Brain Sci.* 1978;4:568-570.
- DE WAAL F. *Bien natural Los orígenes del bien y del mal en los humanos y otros animales.* Barcelona: Herder; 1997.
- DE WAAL F. On the possibility of animal empathy. En: Manstead T., Fridja N. y Fischer A., editors. *Feeling and Emotions: The Amsterdam Symposium,* Cambridge: Cambridge University Press; 2003. p. 379-399.
- DE WAAL F. *Primates and Philosophers: How Morality Evolved.* En: Macedo S, Ober J, editores. United Kingdom: Princeton University Press; 2006.
- DUNBAR, RIM. The social brain hypothesis. *Evol Anthropol.* 1998;178-190.
- FRITH U, MORTON J, LESLIE AM. The cognitive basis of a biological disorder: autism. *Trends Neurosci.* 1991;14:109-128.
- FONAGY P, GERGELY G, JURIST E, TARGET M, editores. *Affect Regulation, Mentalization and the Development of the Self.* Londres: Other Press; 2002. p. 64.
- GORDON R. Folk Psychology as Simulation. En: Davies M, Stone T, editores. *Folk Psychology. The Theory of Mind Debate.* Oxford: Blackwell Publishers; 1995a.
- GORDON R. The Simulation Theory: Objections and Misconceptions. En: Davies M, Stone T, editores. *Folk Psychology. The Theory of Mind Debate.* Oxford: Blackwell Publishers; 1995b.
- HAMILTON WD. The genetical evolution of social behaviour. *J Theor Biol.* 1964;7(1):1-16.
- HAMMOCK EAD, YOUNG LJ. Microsatellite instability generates diversity in brain and sociobehavioral traits. *Science.* 2005;308:1630-1634.
- HARTFIELD E, CACIOPPO TT, RASPÓN RL. Emotional contagion. *Curr Dir Psychol Sci.* 1993;2:96-99.
- HAUSER M. *Moral minds. How nature designed our universal sense of right and wrong,* Nueva York: Harper Collins; 2006. p. 207-217.
- HUMPHREY NK. The social function of intellect. En: Bateson PPG, Hinde RA, editores. *Growing points in ethology,* Cambridge: Cambridge University Press; 1976.
- JOHNSON DB. Altruistic behaviour and the development of self in infants. *Merrill Palmer Q.* 1982;28:379-388.
- KUROSHIMA H, FUJITA K, ADACHI I, IWATA K, FUYUKI A. A capuchin monkey (*cebus apella*) recognizes when people do and do not know the location of food. *Anim Cogn.* 2003;6: 283-291.
- KORSGAARD C. Morality and the distinctiveness of human action. En: De Waal F, Macedo S, Ober J, editores. *Primates and Philosophers: How Morality Evolved by Franz de Waal.* United Kingdom: Princeton University Press; 2006.
- LANGDON R, MICHIE PT, WARD PB, MCCONAGHY CSV, COLTHEART M. Defective self and/or other mentalizing in schizophrenia: a cognitive neuropsychological approach. *Cogn Neuropsychiatry.* 1997;2:167-193.

LESLIE AM. Pretence and representation in infancy: the origins of theory of mind. *Psychol Rev.* 1987;94:84-106.

MOORE C, MACGILLIVRAY S. Altruism, prudence and theory of mind in preschoolers. *New Dir Child Adolesc Dev.* 2004;103;51-62.

OBIOLS J, POUSA E. La teoría de la mente como módulo cerebral evolutivo. En: Sanjuán J, Cela-Conde C.J. *La profecía de Darwin. Del origen de la mente a la psicopatología*, Barcelona: Ars Médica; 2005.

PERNER J. *Understanding the Representational Mind*. Cambridge: MIT Press; 1991.

PREMACK DG, WOODRUFF G. Does the chimpanzee have a theory of mind? *Behav Brain Sci.* 1978;1:515-526.

POVINELLI DJ. Monkeys, apes, mirrors and minds: The evolution of self-awareness in primates. *Hum Evol.* 1987;2:493-507.

POVINELLI DJ, NELSON KE, BOYSEN ST. Inferences about guessing and knowing by chimpanzees (pan troglodytes). *J Comp Psychol.* 1990;104:203-210.

POVINELLI DJ, NELSON KE, BOYSEN ST. Comprehension of special role reversion by chimpanzees: Evidence of empathy? *Anim Behav.* 1992;4:633-640.

POVINELLI DJ. How to create self-recognizing gorillas (but don't try it on macaques). En: Parker ST, Mitchell RW, Boccia ML, editores. *Self-awareness in animals and humans: Developmental perspectives*. Cambridge: Cambridge University Press; 1994.

POVINELLI DJ, PREUSS TM. Theory of mind: evolutionary history of a cognitive specialization. *Trends Neurosci.* 1995;18:418-424.

QUINTANILLA P. Comprensión, imaginación y transformación. *Areté.* 2008;20(1):111-136.

ROSAS A. La evolución de la cooperación. En: Hoyos LE, editor. *Estudios de filosofía política*. Bogotá: Universidad Externado, Universidad Nacional de Colombia; 2004. p. 353.

SINGER P. *The Expanding Circle*. Oxford: Clarendon Press; 1981.

SOBER E, WILSON DS. *Unto others: The Evolution of Psychology of Unselfish Behaviour*. Cambridge MA: Harvard University Press; 1998.

THOMPSON C, BARRESI J, MOORE C. The development of future-oriented prudence and altruism in preschoolers. *Cogn Dev.* 1997;12:199-212.

TRIVERS R. The evolution of reciprocal altruism. *Q Rev Biol.* 1971;46:35-57.

WISPÉ L. History of the Concept of Empathy. En: Eisenberg N, Strayer J, editores. *Empathy and its Development*, Cambridge: Cambridge University Press; 1987. p. 17-37.

WISPÉ L. *The Psychology of Empathy*, Nueva York: Plenum Press; 1991.