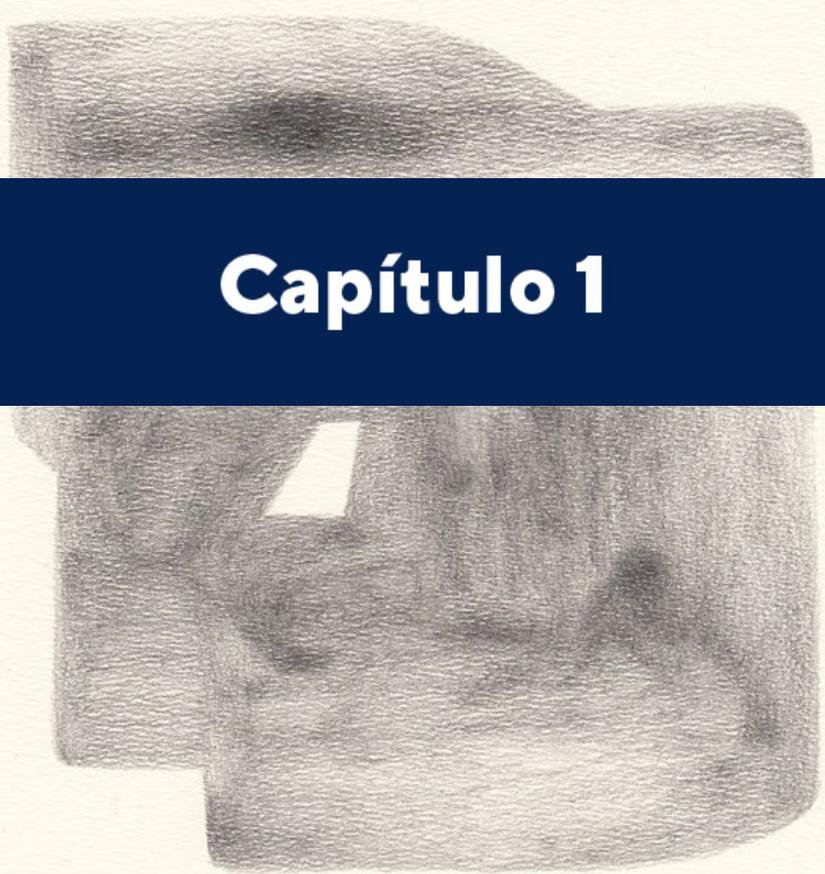


PABLO QUINTANILLA | CARLA MANTILLA | PAOLA CÉPEDA
(editores)

COGNICIÓN SOCIAL Y LENGUAJE

La intersubjetividad en la evolución de la especie
y en el desarrollo del niño



Capítulo 1



FONDO
EDITORIAL

PONTIFICIA UNIVERSIDAD CATÓLICA DEL PERÚ

Cognición social y lenguaje

La intersubjetividad en la evolución de la especie y en el desarrollo del niño

Pablo Quintanilla, Carla Mantilla y Paola Céspedes (editores)

© Pablo Quintanilla, Carla Mantilla y Paola Céspedes, 2014

© Fondo Editorial de la Pontificia Universidad Católica del Perú, 2014

Av. Universitaria 1801, Lima 32, Perú

Teléfono: (51 1) 626-2650

Fax: (51 1) 626-2913

feditor@pucp.edu.pe

www.pucp.edu.pe/publicaciones

Diseño, diagramación, corrección de estilo
y cuidado de la edición: Fondo Editorial PUCP

Primera edición: mayo de 2014

Tiraje: 500 ejemplares

Prohibida la reproducción de este libro por cualquier medio, total o parcialmente,
sin permiso expreso de los editores.

Hecho el Depósito Legal en la Biblioteca Nacional del Perú N° 2014-07083

ISBN: 978-612-4146-80-0

Registro del Proyecto Editorial: 11501361400359

Impreso en Tarea Asociación Gráfica Educativa

Pasaje María Auxiliadora 156, Lima 5, Perú

EVOLUCIÓN Y DESARROLLO DE LA COGNICIÓN SOCIAL

Mente y Lenguaje

Grupo Interdisciplinario de Investigación

Andrés Abugattás, Ricardo Braun, Paola Céspedes, César Escajadillo,
María de los Ángeles Fernández Flecha, Marcos Herrera Burstein,
Carla Mantilla, Luis Manuel Olguín, Jorge Iván Pérez Silva,
Pablo Quintanilla y Carolina Romero

La cognición social, en un sentido amplio, es un concepto que se refiere a todas las funciones mentales implicadas en la comprensión del mundo social; de este modo, es un campo que interesa de manera especial a la psicología social. Esta rama de la psicología ha sido, por algún tiempo, una ciencia cognitiva; sin embargo, luego se vio profundamente influida por la revolución cognitiva que tuvo lugar entre los años 1960 y 1970. A partir de ese momento, la cognición social tomó prestadas las teorías y técnicas de la nueva psicología cognitiva para aplicarlas a los cuestionamientos tradicionales de la psicología social, como las actitudes, la atribución causal, la inferencia social, entre otros. La revolución mencionada se caracterizó, principalmente, por la metáfora del procesamiento de la información, en la cual se plantea la existencia de una serie de procesos mentales por los que circularía la información y daría lugar a la cognición. El circuito sería transitado ante cualquier tipo de cognición, siendo una de ellas la cognición social. Es decir, mientras que los contenidos varían, los procesos serían siempre los mismos. Asimismo, el énfasis fue puesto en las representaciones mentales, con lo cual se dejaba de lado las motivaciones, las emociones, el comportamiento y la interacción social, temas de interés central para los psicólogos sociales. Además, la idea de que todo tipo de cognición tuviera el mismo proceso, sin variación alguna, resultaba un poco radical. Es así como, más tarde, la cognición social se fue ubicando en un punto más intermedio; es decir, si bien incluye la metáfora de procesamiento de información y considera que existen procesos cognitivos generales, también plantea que hay procesos específicos para ciertas áreas. Las teorías modulares se ubicarían aquí (Gilbert, 1999).

A continuación explicaremos la cognición social a partir de la metáfora del procesamiento de la información —la cual es ampliamente aceptada entre los psicólogos sociales—. Esto es, cómo la información es incorporada, elaborada, representada y guardada en la memoria, para luego ser utilizada en la percepción e interacción con el mundo social (Hamilton, 2005). La idea general es que el individuo adquiere nueva información del contexto social, la procesa de distintas maneras, la integra con conocimientos previos y la utiliza para entender a los otros, tomar decisiones y guiar su comportamiento.

Lo primero necesario para que se lleve a cabo este proceso es la atención, la cual permitirá luego la codificación. Usualmente el entorno contiene más información de la que un individuo es capaz de atender; de este modo, la persona se ve forzada a seleccionar ciertos elementos del contexto mientras que otros quedan sin registrar. La atención selectiva es funcional en tanto permite dirigir la atención hacia los elementos más relevantes; sin embargo, al mismo tiempo deja fuera información que podría ser importante para una mayor comprensión de una situación social determinada.

Después de seleccionar y registrar la información, el individuo realiza una serie de elaboraciones sobre aquel material —antes de almacenarlo en la memoria— para comprenderlo mejor y darle un sentido, lo cual a su vez le permite adaptarse a su medio social. Dos tipos de elaboración importantes son la interpretación y la inferencia. Dentro de los tipos de inferencia posibles, el proceso de atribución es uno esencial en tanto nos proporciona la explicación del comportamiento del otro, es decir, sus causas. Ello no solo nos ayuda a entender el porqué de los eventos sino que también nos ayuda a anticipar eventos futuros similares. Un tercer tipo, también importante, es la evaluación, que permite ubicar los eventos en categorías como bueno o malo, bonito o feo, admirable o despreciable, etcétera.

El siguiente paso después de la elaboración consiste en representar la información en la memoria y almacenarla para hacer uso de ella en el futuro. Existen distintas teorías sobre cómo se guarda la información. Algunas sugieren que cada elemento de la información es tomado como un ejemplar y almacenado de manera individual (teorías de ejemplares); otras señalan que los elementos se almacenan interconectados, formando circuitos (teorías de circuitos); y otras indican que de los elementos de la información se obtiene un significado, un concepto más general —esquema o prototipo— bajo el cual se los agrupa (teorías basadas en la abstracción). Independientemente de cuál sea la teoría más adecuada para entender el almacenamiento de la información, lo cierto es que aquella que se almacena no consiste en una representación objetiva de la realidad. Primero, se han seleccionado únicamente ciertos elementos y, en segundo lugar, ha habido una serie de procesos elaborativos previos al almacenamiento, lo que implica una transformación de la información.

Y lo importante es que esa información transformada es la que va a guiar la acción futura del individuo.

Luego de almacenada la información, esta puede ser recuperada (evocada) para formar juicios y guiar el comportamiento. Sin embargo, así como el almacenamiento no es objetivo, tampoco lo es la recuperación de la información. Este proceso va a estar influido por las metas o propósitos inmediatos de la persona, seleccionando ciertos elementos y no otros. Se pierde la mirada global. Algunos elementos serán recordados más fácilmente que otros por el impacto que tuvieron en el individuo y podrán evocar otros recuerdos que tengan alguna conexión con los primeros. Es más, nuestros objetivos actuales pueden reconstruir nuestros recuerdos de tal modo que encajen con nuestras creencias y metas actuales.

Finalmente, como resultado de todo este proceso se da la respuesta, que puede incluir juicios, reacciones afectivas, decisiones y comportamientos. Como hemos visto, una serie de factores influyen en el procesamiento de la información y lo convierte en un proceso único y capaz de variar de acuerdo al momento en que se realiza. De este modo, una misma persona puede parecer inconsistente en tanto frente a estímulos similares puede realizar procesamientos distintos y, por tanto, mostrar respuestas diferentes.

Algo sumamente interesante que se ha observado en varias investigaciones es que, a menudo, llevamos a cabo el procesamiento de información social de manera automática, no deliberada. Es decir, muchas veces no nos damos cuenta de aquello que estamos registrando, elaborando y almacenando en la mente, pues ni siquiera interrumpe las tareas cognitivas que se estén realizando de manera paralela y consciente. No obstante, se ha encontrado que este procesamiento automático tiene gran influencia en nuestras respuestas y, por tanto, en la adaptación a nuestro entorno social.

Hasta aquí se ha descrito, fundamentalmente, los pasos del procesamiento de la información social, mas no se ha dicho nada sobre los contenidos. De manera general, nos manejamos con estructuras cognitivas que son conceptos que se han ido construyendo a partir de las propias experiencias y que representan nuestro conocimiento, nuestras creencias y expectativas en distintas áreas de nuestra vida. Estas estructuras, como las teorías implícitas de la personalidad o los estereotipos, pueden influir en cada paso del procesamiento de la información descrito más arriba. Ayudan a mantener el *statu quo* y nos permiten anticipar el futuro. Esto debido a que, en la mayoría de los casos, la nueva información es integrada a las estructuras cognitivas existentes (Hamilton, 2005).

Varias de estas estructuras cognitivas funcionan como atajos mentales que nos ayudan a realizar un menor gasto de energía al procesar la información nueva. Sin embargo, utilizarlos conlleva el riesgo de caer en error. Estos se denominan *heurísticos*.

Por ejemplo, en lugar de hacer un análisis detallado de una persona la ubicamos en una categoría por la similitud que le encontramos con algún grupo particular (persona de barba y pelo largo, con una vestimenta desgarbada y sandalias nos podría remitir a un *hippie*). Otro heurístico común es el que nos permite pensar en la probabilidad de un atributo a partir de la facilidad con que recordamos ese atributo al pensar, por ejemplo, en un grupo (si cuando pensamos en gente negra lo primero que se nos viene a la mente es que cantan muy bien, cuando estemos frente a una persona negra le vamos a atribuir esa cualidad casi de inmediato).

De otro lado, es interesante mencionar que nuestras propias metas y objetivos pueden influir de manera importante en el procesamiento de información social y en la respuesta que a partir de él se genere. Esto es, no solo puede determinar qué se registra sino cómo se registra. Por ejemplo, se ha encontrado que las personas tienden a ser menos exigentes en cuanto a las evidencias cuando una propuesta es cercana a nuestro propio pensamiento. En contraste, si se aleja de él, es más probable que seamos más meticulosos al momento de evaluar la información para constatar su validez.

A partir de lo expuesto anteriormente vemos cómo la cognición social, aunque presentada desde un modelo que la describe y explica de una manera relativamente sencilla, es en realidad un proceso sumamente complejo en el que intervienen innumerables variables de distinto tipo. A nosotros nos interesa centrarnos especialmente en una sección del procesamiento de información que se encuentra dentro de las actividades de elaboración: la atribución.

Los procesos de atribución no solo han sido de interés para la psicología social sino también para otras áreas de estudio, utilizando sus propios métodos de investigación y conceptos particulares. En las ciencias cognitivas, y por aporte de filósofos, por ejemplo, un concepto particularmente relevante en este contexto es el de metarrepresentación. La atribución de representaciones mentales es, pues, una forma de representación de segundo grado: es una representación de una representación. Cuando intentamos explicar o predecir el comportamiento de algún agente solemos emplear ciertos verbos que describen la actividad mental que, asumimos, acompaña al comportamiento. Así, decimos que el agente actuó de cierta manera porque quería o deseaba algo, porque pensaba en algo, etcétera. Querer, desear o pensar en algo implica poder dirigir la conciencia hacia aquello que nos ocupa, lo cual implica, a su vez, alguna forma de representación.

La metarrepresentación adopta dos formas básicas. Cuando la empleamos en agentes distintos de nosotros mismos recibe el nombre de 'lectura de mentes' (*mindreading*). Otra manera de referirse a ella es como atribución de representaciones en tercera persona. Pero, también podemos realizar atribuciones de representaciones en primera persona, es decir, vernos a nosotros mismos como agentes. En este caso, se habla de *metacognición*.

La metarrepresentación es una poderosa herramienta de la sociabilidad humana. Debido a su poder predictivo es comparable con la postulación de hipótesis empíricas de la actividad científica. En ciencias cognitivas se habla de ‘teoría de la mente’ (*theory of mind*) o de *psicología folk* para hacer referencia al conocimiento común, prácticamente intuitivo, que poseemos sobre la relación entre los contenidos mentales y el comportamiento de los agentes, incluyéndonos a nosotros mismos. El término *teorías de la teoría de la mente* (TT) se refiere a los posibles modelos teóricos que explican el conocimiento psicológico folk.

Muchas veces quienes comparan nuestras labores cotidianas de intérpretes psicológicos con el empleo de teorías científicas, sostienen que nuestras predicciones respecto del comportamiento de los agentes son el producto de deducciones hechas a partir de principios psicológicos generales, ya sean implícitos o explícitos. Sin embargo, también hay quienes consideran que para explicar la interpretación psicológica no es necesario postular una suerte de conocimiento teórico folk. La alternativa más conocida a las TT son las llamadas *teorías de la simulación* (TS). Las teorías de la simulación se limitan a postular que todo lo que necesitamos para predecir el comportamiento de otros agentes es intentar proyectarnos a su situación y, desde ahí, deducir qué podrían estar pensando. En otras palabras, en tanto intérpretes del comportamiento, simulamos ser los agentes cuyas acciones queremos explicar o predecir.

Más allá de las evidentes discrepancias generales, un punto importante en el debate entre defensores de las TT y defensores de las TS es en torno de la atribución de representaciones en primera persona. Para el segundo grupo de autores la interpretación de comportamiento de los agentes parte de la aprehensión directa de los propios estados mentales del intérprete. Como dijimos arriba, todo el conocimiento que podemos aspirar a tener de las mentes de los demás es solo una proyección de lo que hemos simulado, en nosotros mismos, como la situación del otro. Una consecuencia de esta tesis es que la metacognición es considerablemente menos falible que la lectura de mentes, pues solo el conocimiento de nuestras propias mentes sería directo. Se sostiene, además, que la seguridad superior de la metacognición, también conocida como la *autoridad de la primera persona*, se debe a que poseemos un mecanismo de introspección mental. Uno de los defensores más conocidos de esta posición es Alvin Goldman (2006). Algunos simpatizantes de las TT niegan la necesidad de postular un mecanismo semejante y, por ello, admiten que la metacognición puede ser casi tan falible como la lectura de mentes. Como veremos más adelante, autores como Peter Carruthers (2009), conocido defensor de las TT, piensan que el debate debe solucionarse mediante la comparación de las explicaciones filogenéticas de los mecanismos en cuestión.

Como podemos apreciar, las perspectivas hasta ahora expuestas se centran, esencialmente, en los procesos cognitivos implicados en la construcción de representaciones mentales sobre los otros y sobre nosotros mismos, lo cual es fundamental para la cognición social. Sin embargo, existe otro grupo de teorías que más bien se concentra en los procesos sociales que permitirían aflorar aquellas capacidades mentales necesarias para la cognición social. Aclaramos que ninguno de los grupos niega la importancia social o la importancia del factor biológico, respectivamente; se trata de distintos enfoques en la investigación de la cognición social.

Dentro de las teorías relacionales encontramos la de la mentalización que, a su vez, se apoya en la teoría del apego. La capacidad para la mentalización involucra, en realidad, una serie de habilidades que permiten al individuo interpretar su propia conducta y la de los demás a partir de la atribución de estados mentales —como sentimientos, deseos, creencias, entre otros (Fonagy, 2008)—. Para que la mentalización se consolide de manera adecuada, es necesario que se genere un apego seguro en el niño y, para que eso ocurra, los cuidadores deben ser capaces de leer las señales del bebé y responder a ellas de manera adecuada. Algunas de las labores más importantes de los adultos son amplificar las expresiones emocionales positivas del bebé y modular las negativas. De manera general, es a través del intercambio con los padres o cuidadores que el infante hace uso de las funciones ya maduras del cerebro adulto para organizar sus propios estados mentales. Es gracias a esos encuentros interpersonales que el bebé logra desarrollar una serie de funciones mentales muy importantes como la memoria, la narración, las emociones, las representaciones y los estados mentales (Siegel, 1999). Todas ellas esenciales para un adecuado desarrollo y desenvolvimiento en el mundo social.

Este primer capítulo del trabajo está constituido por dos subtítulos. El primero se centrará en algunas teorías sobre la historia filogenética de las facultades cognitivas humanas implicadas en la cognición social y, el segundo, abordará el tema desde una perspectiva del desarrollo, es decir, desde una mirada ontogenética, e incluirá algunos planteamientos de las teorías relacionales.

1. LA EVOLUCIÓN DE LA COGNICIÓN SOCIAL

Este apartado cuenta con cinco partes principales. La primera es una breve introducción histórica a la investigación filogenética de la cognición social que, como en muchas otras formas de estudios filogenéticos, tiene sus raíces en Darwin. La segunda parte está dedicada a algunas teorías sobre la evolución de los elementos del cerebro humano que influyen en la interacción social. En la tercera parte, veremos discusiones actuales sobre conceptos centrales para las teorías filogenéticas de la atribución psicológica. La cuarta parte trata sobre la hipótesis de la inteligencia social, la explicación

más aceptada de la evolución de la atribución psicológica. En la última parte del apartado, veremos algunos problemas filosóficos que hoy se abordan empleando argumentos extraídos de las teorías sobre la evolución de la cognición social.

1.1. La teoría darwiniana de la selección natural de las facultades sociales humanas

Si bien la idea de que las especies evolucionan o han evolucionado se remonta hasta prácticamente los orígenes de la filosofía griega, con Anaximandro, fue a partir de la postulación de la teoría de la selección natural que las investigaciones evolutivas han sido aprovechadas como una importante herramienta para la discusión en diversas ciencias. Como se sabe, Darwin sostuvo que los rasgos biológicos aleatoriamente heredados son seleccionados de acuerdo con un principio simple, a saber, si son útiles para la supervivencia de los individuos. Naturalmente, los rasgos más exitosos son los que se transmiten sistemáticamente de generación en generación, pues gracias a ellos los organismos portadores tienden a sobrevivir y reproducirse con eficacia.

Uno de los méritos de Darwin es el haber ofrecido un argumento, probablemente definitivo, contra el esencialismo platónico y aristotélico. Para Aristóteles, las especies animales son eternas e inmutables. Los individuos participan de una esencia común que se puede rastrear hasta Platón, de hecho, la palabra *especie* fue una traducción latina para *forma* o *eidos*. Darwin rechaza la inmutabilidad de las especies, para él las especies están siempre evolucionando, en el sentido de transformándose (de hecho, antes que se usara la palabra *evolución* se empleaba *transformismo*), y si bien los individuos de la misma especie comparten muchos rasgos, no hay entre ellos rasgos esenciales y accidentales, sino que todos los rasgos son accidentales y llamamos *esenciales* a aquellos rasgos accidentales compartidos por muchos individuos en un momento dado.

Según Darwin, entonces, las especies no son entidades fijas, como sostenía el llamado fijismo de su época que, siguiendo a Aristóteles, defendía que los individuos que pertenecen a una especie comparten una esencia común, que siendo estable permite que un individuo sea lo que es y no otro. Asimismo, habría una diferencia cualitativa y no meramente cuantitativa entre especies diferentes. Para Darwin, por el contrario, no solo no hay esencias estables compartidas por los individuos que pertenecen a una especie sino que, además, las diferencias entre especies son meramente cuantitativas. Es verdad que los individuos de una especie pueden reproducirse entre sí y no con miembros de otras especies (la posibilidad de la reproducción suele considerarse el criterio principal para distinguir entre especies), pero eso solo se produce cuando dos grupos de individuos se han apartado genéticamente lo suficiente entre sí para que sea imposible que se reproduzcan, que es precisamente lo que se llama el proceso de *especiación* o formación de las especies.

Fue el mismo Darwin quien no tardó en explotar la potencia explicativa de su propia teoría, empleándola no solo en la investigación de los orígenes de rasgos meramente fisiológicos de nuestra especie. No es extraño, entonces, que los capítulos III, IV y V de *El origen del hombre* (1969) estén dedicados a defender la tesis de que las diferencias entre las facultades mentales en general, y morales en particular, del ser humano y de los animales no son esenciales. Por el contrario, en muchos casos se trataría de distinciones de grado determinadas, evidentemente, por la selección natural.

El primer problema que debe enfrentar Darwin para defender su tesis consiste en la distinción entre lo perteneciente al mero instinto y la inteligencia, entendida esta, a veces, como el libre albedrío y, quizás más importante aún, también como lo aprendido y no fijo (1969, p. 87). Como se verá más adelante, a medida que la investigación avanza, la distinción se hace más opaca, lo cual era de esperarse. Como advierte el autor desde un principio, una vez asumida la distinción no es fácil determinar qué facultades pertenecen a una y otra categoría. Podría ser, por ejemplo, que aquello que tuvo origen gracias al uso de la inteligencia se convierta, a través de las generaciones, en actos instintivos. Además, no siempre se cumple la idea clásica de que la relación entre instinto e inteligencia es inversamente proporcional (pp. 86-87).

No obstante lo dicho arriba y aunque sus argumentos no carezcan de tensiones, Darwin parece hacer recaer el peso de la distinción en cuestión en una mayor *automaticidad* de los actos animales en comparación con aquellos de los humanos, provistos de conciencia. Afirma que el fundamento del comportamiento humano radica en «la memoria de los acontecimientos pasados, en la previsión, en la razón y en la imaginación» (1969, p. 89). Y aunque en algunos animales podemos ver acciones que se asemejan en muchos aspectos de su ejecución y fines a las humanas, en aquellas «la variabilidad de los órganos mentales, unida a la selección natural, es la que adquiere paso a paso capacidad para la ejecución de tales acciones sin que contribuyese a ello la inteligencia ni la conciencia del animal en cada una de las distintas generaciones» (p. 89). El mismo punto es reforzado cuando el autor hace referencia a la adquisición de ciertas habilidades en los animales: en muchos casos, para estos bastaría con la sencilla imitación, o a veces ni siquiera eso, para que aquellas alcancen su máximo nivel de desarrollo, mientras que tratándose de los individuos humanos la experiencia juega un papel preponderante.

Frente a la carencia de información sobre la evolución y el desarrollo de las diversas facultades mentales tanto en animales como en humanos, el principal recurso argumentativo de Darwin consiste en citar muchos ejemplos de investigaciones de naturalistas de la época, incluidas algunas suyas, e intentar mostrar cómo ciertos comportamientos animales registrados son indicadores de emociones y procesos mentales semejantes a los nuestros. Resultan particularmente notables las observaciones que

el autor hace acerca de la capacidad que algunos animales poseen para resolver tareas complejas o que distan considerablemente de los procesos que usualmente resuelven mediante hábitos. En muchos casos, ante el error (generalmente repetido, aunque en los mamíferos superiores esto no es necesario), aquellos pueden corregir el curso de sus acciones y realizarlas exitosamente. Sin mayores problemas, Darwin compara esta facultad animal con lo que normalmente llamamos *reflexión, deliberación* o simplemente *raciocinio* en los seres humanos, salvando las distancias respecto de nuestra muchísimo mayor capacidad de observación y deducción.

De la misma manera, Darwin considera al fenómeno del lenguaje. La facultad de asociar sonidos a conceptos más o menos bien determinados no es un problema tratándose de animales; bastaría con observar cómo estos responden a ciertas voces nuestras. Inclusive, el autor llega a poner en cuestión que el encadenamiento de ideas requiera necesariamente del empleo de signos lingüísticos (1969, p. 117), lo cual también echa algunas luces sobre la teoría lingüística que aquel aceptaba. Tampoco sería un problema la capacidad de articular signos, aunque en torno de este punto el naturalista emplea el cuestionable ejemplo de las aves capaces de imitarnos acústicamente (p. 113). Nuevamente, la distinción debe ser una cuestión de grado, y en este caso particular ocurre que los seres humanos simplemente poseemos una muchísima mayor capacidad de enlazar libremente signos y significados (p. 114).

Darwin también se pone de lado de quienes consideran que, en el fondo, es el comportamiento moral lo que marca la principal diferencia entre seres humanos y animales. No obstante, según pensaba, la moral también puede explicarse genealógicamente empleando la teoría de la selección natural. Es más, si los antepasados de los hombres (y otras especies no tan cercanas, como muchos mamíferos) ya contaban con formas básicas de simpatía¹ y de ciertos instintos de conservación grupal, bastaría con añadir a la ecuación una capacidad de raciocinio semejante a la nuestra y obtendríamos, de forma natural, el sentido moral humano tal y como lo entendemos hoy en día.

A pesar de las dificultades que el propio Darwin reconoce cuando se intenta atribuir a los animales estados mentales y afectivos que sin dificultad alguna atribuimos a otras personas, no duda en responsabilizar a la simpatía por muchos actos altruistas que se observan en el mundo natural². Igualmente, piensa que es difícil explicar

¹ Darwin no distingue aún entre *simpatía* y *empatía*, aunque podemos atrevernos a afirmar, a juzgar por los ejemplos que utiliza, que lo que él llama *simpatía* se acerca mucho más a lo que los estudiosos contemporáneos llamarían por el mismo nombre.

² Como veremos más adelante, a Darwin no le caben dudas sobre la preponderancia de la simpatía en el mundo humano. Basta con leer la siguiente cita: «Cuanto más importancia ponga el hombre en el aprecio de los demás, mayor fuerza tendrá en él el sentimiento innato o adquirido de simpatía, y asimismo su propia capacidad para deducir las consecuencias de sus actos» (1969, p. 153).

comportamientos tales como las rondas de vigilancia y otro tipo de precauciones en diversas especies animales sin apelar a un instinto común de conservación de grupo. Resulta interesante señalar, a propósito de la simpatía y otros instintos morales, que tomarlos como punto de partida para la evolución de la moral permite evitar tesis simplistas que apelan a relaciones de placer y dolor como la fuente original del comportamiento social. La simpatía y el instinto de conservación de grupo conducen a los individuos a acciones que incluso podrían poner en riesgo su propia seguridad, y esta tendencia parece haberse mantenido y alimentado en el caso de los seres humanos, lo cual confirmaría la hipótesis darwiniana de que la selección natural ha favorecido el comportamiento altruista a través de las eras y las especies.

Antes dijimos que las facultades mentales superiores de los seres humanos jugaban un rol determinante en el distanciamiento entre estos y sus antepasados animales, a pesar de la relación de parentesco. Siguiendo algunas pistas que deja Darwin sobre este punto, podemos inferir lo que sigue. Tratándose de los animales, los mecanismos que configuran su sociabilidad están más fuertemente determinados por los instintos. Si se quiere, podemos decir, utilizando lenguaje informático, que están *preprogramados* en un porcentaje mucho mayor que los mecanismos humanos equivalentes. En este último caso, la carencia de instintos es subsanada con la mayor potencia del raciocinio (1969, p. 146) (podemos agregar, también, las mayores posibilidades de comunicación que nos brinda nuestro lenguaje). Una de las consecuencias de ello es la importancia enorme que poseen las opiniones de la comunidad cuando nos disponemos a actuar. De hecho, Darwin piensa que nuestra facultad de considerar los juicios que los demás pueden hacer sobre nuestros actos sigue evolucionando a medida que nos volvemos más apegados a formas cada vez más abstractas de valoración de los primeros y se atenúa la relevancia de las impresiones subsecuentes, ya sean de placer o de dolor (p. 147).

Si bien, como vimos, las facultades racionales parecen ser un elemento fundamental de la distinción entre moral humana y animal, no por ello se debe disminuir el valor del instinto en los asuntos morales, ni siquiera tratándose del ser humano. Ya hemos visto que Darwin parte en su investigación de la diferenciación entre razón e instinto, mas esta tampoco es clara tratándose de lo moral. De una parte, sabemos que la acción de lo que podemos llamar «instinto» en los asuntos morales humanos no es poca. De otra, también en este ámbito encontramos cierta deliberación entre los animales, a quienes no pocas veces puede vérselos indecisos ante más de una opción para resolver una situación difícil. Tanto en los seres humanos como en los animales que viven en sociedad, la reflexión o deliberación moral proviene de la constante presencia en los individuos de «impresiones o imágenes del pasado» (1969, p. 150). Pero, dadas las mayores facultades mentales humanas, dichas impresiones,

que incluyen los juicios aprobatorios o desaprobatorios de nuestros compañeros en la sociedad, son muchísimo más persistentes (persistencia que, sobre todo en lo que respecta a los juicios de los demás, se apoya en la simpatía). El resultado de su persistencia son los sentimientos como el arrepentimiento, la vergüenza o el remordimiento. A su vez, dichos sentimientos son distintas caras de lo que usualmente llamamos *conciencia moral*, y es su peso lo que nos obliga a revisar nuestras acciones y querer corregirlas en las siguientes ocasiones (p. 152).

La simpatía, entonces, se ha visto favorecida por la selección natural debido a su capacidad de fomentar la sociabilidad de humanos y ciertos animales. De ahí que Darwin afirme que es la consecución del *bien general*³ lo que ha servido como principio regulador de la evolución de la moral humana, en la cual la preponderancia de la simpatía es mucho mayor que en otras especies sociales. Es más, el autor no duda en afirmar que, entre nosotros, la búsqueda de la felicidad queda constantemente relegada a un segundo plano con relación al bien de todos. En virtud del éxito que este principio demuestra en la formación de sociedades cada vez más grandes y complejas a lo largo de la historia humana, fuerza la expansión de la simpatía entre los individuos más allá de los límites de la propia comunidad, de manera que los seres humanos poseemos, hoy en día, la capacidad de sistemáticamente sentir simpatía hacia cualquier miembro de nuestra especie e, inclusive, con miembros de otras especies⁴.

A pesar de las tensiones manifiestas del texto, de que en la época recién se desarrollaban las primeras investigaciones al respecto y de que, además, muchos de los ejemplos y referencias citadas por el autor hoy en día serían altamente cuestionables, encontramos entre sus páginas las raíces de muchas teorías actuales o, en algunos casos, intuiciones que recientemente han tomado vigor entre algunos investigadores. Desde ya, el intento de Darwin por encontrar en formas de vida no humanas facultades mentales similares, en alguna medida, a las humanas lo conduce a poner en tela de juicio (aunque algunas veces contra sus propias premisas) ideas como la antes mencionada separación clara y distinta entre la razón y el instinto. Ciertamente, la distinción común en la filosofía de la mente contemporánea entre *Sistema 1* y *Sistema 2* ha ampliado el rango de funciones que estamos dispuestos

³ Darwin define el término de la siguiente manera: «[...] es el medio de producir, dentro de las condiciones existentes, el mayor número de individuos en pleno vigor y plena salud, dotados de facultades todo lo perfectas posible» (1969, p. 162). En vista de ello, no es de extrañar que el propio Darwin considere que su tesis constituye una refutación de cualquier explicación moral que sitúe en el punto de partida principios derivados del egoísmo.

⁴ Ya antes David Hume había descrito, en el tercer libro de su obra *A Treatise of Human Nature*, un movimiento similar de los lazos morales entre los individuos humanos, desde niveles más básicos y locales, hasta niveles más generales o abstractos (cabe recordar la cercanía entre Hume y Adam Smith, quien es citado por Darwin en la obra que aquí tratamos).

a aceptar bajo la categoría de *razonamiento*. Muchas veces, se define lo perteneciente al Sistema 1 de una manera similar a la que Darwin pensaba lo instintivo, como procesos, dijimos, más automáticos y con una mayor carga hereditaria, mientras que el Sistema 2 se acerca a lo que anteriormente se pensaba como lo propiamente racional (procesos más complejos, conscientes y culturalmente variables). Esta nueva distinción, a su vez, se vuelve a poner en cuestión hoy en día (Carruthers, 2009).

Respecto de muchas investigaciones contemporáneas en torno de los orígenes de la moral, también podemos señalar huellas darwinianas en algunas ideas importantes. Ya que Darwin considera que la sociabilidad es el principal factor que ha determinado y determina la evolución de nuestra especie, se afirma que él es también el padre de la teoría de la selección de grupos. Esto se debe a su tesis de que, finalmente, la simpatía, el altruismo y otras facultades de los seres altamente sociables, producen sociedades que prevalecen sobre otras donde las mismas cualidades se encuentran menos desarrolladas entre sus miembros (2009, pp. 427-429)⁵.

1.2. El desarrollo filogenético del cerebro social

Como vimos, Darwin propuso una historia evolutiva naturalista de las facultades sociales humanas. Hoy en día, gracias a los avances en la tecnología, las neurociencias constituyen una nueva entrada para la exploración de dichas facultades a partir del estudio del cerebro social, es decir, investigar aquellas partes del cerebro vinculadas al área social tanto a nivel estructural como funcional. Un estudio profundo de ello permite, además, vislumbrar aquellos sistemas que han evolucionado a lo largo del tiempo debido a las presiones de selección.

Barton (2007) presenta el recorrido por el cual se intentó analizar las diferencias cognitivas entre especies. El punto de inicio fue el estudio del tamaño cerebral total. En efecto, algunas investigaciones encontraron que la evolución del tamaño del cerebro correlaciona con ciertas especializaciones que probablemente responden a una selección de capacidades cognitivas. Sin embargo, otros estudios han obtenido resultados dispares entre sí y existe la posibilidad de que las diferencias en el tamaño de los cerebros se deban a diversas variables.

⁵ Como afirma Quintanilla (2008, pp. 5-6), Darwin no resuelve el problema acerca de cómo se desarrolló la competencia entre los individuos más altruistas y los más egoístas, interrogante que ha dado lugar al desarrollo de teorías como la de la selección de parentesco o del altruismo recíproco. A favor de Darwin podemos agregar que, en el capítulo V de *El origen del hombre* (p. 176), el autor adelanta una hipótesis acerca del motivo del desarrollo de las cualidades altruistas en los individuos: «En primer lugar, aumentándose de día en día las facultades de razonar y prever los sucesos, cada hombre se convenció bien pronto de que, ayudando a sus compañeros, los obligaba a pagarle después en igual moneda sus servicios». Esta afirmación se acerca mucho a la idea del altruismo recíproco.

Debido a ello se propuso otro método, un poco más minucioso, que consiste en investigar las variaciones de ciertos sectores del cerebro de manera independiente al tamaño total del mismo. A esta variación de regiones se le conoce con el nombre de evolución en mosaico y es producto de ciertas presiones de selección. Sin embargo, si bien ciertas subdivisiones se han agrupado y han evolucionado de un modo particular e independiente para cumplir una serie de funciones —como el cerebelo, la amígdala o el hipotálamo—, dentro de ellas existen grandes diferencias. Esto es, en cada subdivisión existen áreas que se encargan de funciones diferentes —por ejemplo, la amígdala se divide en el grupo basolateral, el núcleo centromedial y el núcleo cortical—, lo cual dificulta una comprensión detallada de la evolución de funciones específicas estudiando regiones cerebrales como un todo. Además, y esto es quizá lo más importante, las áreas de las subdivisiones presentan amplias conexiones con muchas otras regiones del cerebro. No se trata, entonces, de circuitos cerrados en cada subdivisión sino más bien de circuitos abiertos que interactúan continuamente con otros sectores.

En consecuencia, para determinar cuáles han sido los procesos cognitivos seleccionados por la evolución es necesario indagar sobre los sistemas neuronales funcionalmente integrados, que incluyen distintas áreas del cerebro como las que procesan información sensorial, motora y emocional. En este sentido, los procesos cognitivos no deben ser tratados como ajenos a estos últimos. Lo que se espera es que los componentes de cada sistema muestren una evolución correlacionada, es decir, que hayan cambiado de tamaño conjuntamente pero independientemente del cambio en otros sistemas.

Un ejemplo de lo anterior es el circuito ventral-prefrontal, en el que se aprecian grandes conexiones entre ciertas áreas del lóbulo temporal, la amígdala y la corteza prefrontal. La región temporal es esencial para la interacción social en tanto responde a caras de manera selectiva; la amígdala, por su parte, también responde selectivamente a caras y a señales emocionales como el temor y la rabia, además de a las expresiones faciales y movimientos corporales; y el área prefrontal se encarga de procesar la información obtenida de ambas fuentes. Así, este circuito procesa señales sociales en un contexto emocional.

Siguiendo con esta línea de pensamiento, Panksepp (2007) sostiene que el cerebro social humano surge de una serie de tendencias emocionales innatas que interactúan con procesos cognitivos de nivel superior. Lo interesante de su propuesta es que estas tendencias constituyen la base sobre la cual los procesos cognitivos superiores se despliegan. Hasta el momento se han identificado siete sistemas emocionales nucleares. Los cuatro primeros se encuentran en todos los vertebrados: el temor, la ira, la búsqueda y el impulso sexual. Ellos constituyen los principales elementos para

los sentimientos de ansiedad, rabia, deseo y erotismo en mamíferos. En estos últimos han surgido, además, los sistemas de cuidado, pánico y juego, que son las bases esenciales para los sentimientos de cuidado y crianza, tristeza y ansiedad de separación, y alegría social, respectivamente.

Estos afectos son *memorias de evolución* que permiten anticipar y afrontar con éxito las situaciones que amenazan o favorecen la supervivencia. No se trata, entonces, de un aprendizaje efectuado durante el ciclo vital sino de un aprendizaje a lo largo de generaciones que ha sido transmitido genéticamente para la subsistencia de ciertas especies que se desenvuelven en ambientes particulares. En humanos, estos afectos son luego analizados a la luz del aprendizaje social, lo cual da lugar a una complejidad emocional que no se encuentra en los animales. Es importante recalcar que no se está sugiriendo que los cerebros tienen contenidos específicos sino disposiciones que serán desarrolladas en el transcurso del desarrollo.

Las investigaciones sugieren que las capacidades socioemocionales más tempranas son la habilidad para sentir ansiedad cuando se detecta la pérdida de apoyo social —y comunicar vocalmente ese estado para incrementar la probabilidad de reunión— y la tendencia a sentir comodidad ante el apoyo social, lo cual se aprecia a través de sonrisas y risas. En edades tempranas, existe la tendencia a generar vínculos positivos a través de, por ejemplo, juegos físicos toscos. Estas emociones esenciales en la juventud —ansiedad de separación y disposición al juego— son fundamentales para la calidad de los vínculos con los cuidadores y, posteriormente, la amistad con otros, lo cual a su vez va sentando las bases para las habilidades sociosexuales. En cambio, la sexualidad podría haber servido como una precondition evolutiva a partir de la cual los impulsos maternos eventualmente surgieron. De este modo, el entendimiento de las relaciones sociales más complejas podría basarse en la comprensión de la ansiedad de separación y la disposición al juego de los más jóvenes, por un lado, y los impulsos sexuales y maternos de los adultos por el otro.

Mientras que Panksepp se concentra en las tendencias emocionales básicas, Jellema y Perrett (2007) se ocupan de algunos de los circuitos neuronales fundamentales para la cognición social, aquellos concernientes al procesamiento visual. Se trata de sistemas que proporcionan la base mecánica para la comprensión de las acciones de los demás, mas no dependen del entendimiento de las mentes de otros. Así, son capaces de vincular las acciones de las demás personas con claves situacionales de tal modo que permiten generar una relación causal y de predecir probables eventos o resultados gracias al acceso de información perceptual inmediatamente anterior. Es a partir de estos sistemas más básicos que se construirán otros de mayor complejidad que permitirán ya no solo predecir la conducta ajena sino explicarla de un modo más sofisticado.

Existen dos corrientes corticales distintas para el procesamiento visual: una que va desde la corteza visual hasta la temporal inferior y otra que también parte de la corteza visual pero se dirige a la parietal inferior. La primera se encarga de registrar la percepción visual y la segunda se ocupa de la acción, ambas haciendo uso de una codificación espacial. Primeramente el cuerpo es dividido en sectores claves como los ojos, la boca, la cabeza, el dedo, la mano y el brazo para luego ser analizados en términos de movimiento. Finalmente, en el surco temporal superior (STS) confluyen ambos circuitos para procesar el comportamiento de otros a partir de posturas y animaciones clave. Por ejemplo, algunas células del STS se activan al observar a un agente realizando el acto de alcanzar siempre y cuando la cabeza y los ojos estén apuntando a la misma dirección, lo cual ayuda a detectar acciones intencionales. Asimismo, otras células se activan frente a actos como recoger, rasgar y manipular.

Sin embargo, esta información necesita unirse a sistemas de mayor complejidad para atribuir estados mentales a los demás. Uno de estos, que se conecta estrechamente con el STS, es el sistema de las neuronas espejo, el cual permite entender el significado de las acciones mediante una simulación interna de las mismas. Las neuronas espejo son un tipo particular de neuronas motoras que se caracterizan por codificar actos motores que contienen una meta; por ejemplo, cargar, agarrar, manipular y rasgar. Como se verá más adelante, muchos consideran que a partir de este sistema es posible descubrir las intenciones de las demás personas mediante un proceso que permite experimentar los actos del otro como si fueran propios. Naturalmente, esta relación entre el STS de las neuronas espejo es explotada por quienes, como Alvin Goldman (2006), sostienen que la interpretación de otros agentes es básicamente un proceso de simulación (como se vio en la introducción a esta sección del trabajo, se trata de las teorías de la simulación o TS).

Este tipo de neuronas fue descubierto en monos macacos primero en el área F5 (una parte de la corteza premotora ventral) y, más adelante, en la parte rostral del lóbulo parietal inferior (IPL). En los humanos, la localización de las neuronas espejo es similar a la de los macacos; se encuentran en el lóbulo parietal inferior y en la parte baja del giro precentral (corteza premotora ventral), además de la parte posterior del giro frontal inferior. Así, las regiones parietal y frontal constituyen el núcleo del sistema de las neuronas espejo (Rizzolatti & Fogassi, 2007).

Dentro de las neuronas del área F5, existe un grupo que responde a estímulos visuales (neuronas visomotoras) el cual, a su vez, se divide en dos categorías: las neuronas canónicas y las neuronas espejo. Las primeras se activan ante objetos tridimensionales mientras que las segundas lo hacen al observar a un agente realizando una acción y cuando uno mismo realiza la misma acción o una similar. De este modo, existe una congruencia entre las respuestas visuales y las respuestas motoras.

El funcionamiento del sistema de neuronas espejo es como sigue: cada individuo tiene una representación de los actos motores, codificado en las cortezas parietal y frontal. Estas representaciones pueden ser evocadas internamente cuando una persona decide realizar un acto. En este caso el agente sabe (predice) el desenlace de la acción. Cuando la misma representación motora es activada al observar a otro individuo realizando ese acto motor, el observador sabe lo que el otro está haciendo porque la observación del acto activa, en su propia corteza parietal y frontal, la representación motora de esa acción. De esta manera, la acción de otro individuo es como si fuera propia y ello permite descubrir la meta detrás del acto (2007).

Dado que las neuronas espejo se activan frente a acciones con objetivos, algunos investigadores se preguntaron si entonces estas debieran activarse cuando el significado del acto fuera comprendido, aún en ausencia de información visual. Los experimentos realizados confirmaron esta hipótesis. En uno de ellos, la parte final de la acción fue ocultada tras una pantalla y, no obstante, las neuronas espejo se activaron. En otro experimento, se les permitió a los monos ver y oír cómo se rompía un pedazo de papel y, en un segundo momento, se los expuso solo al sonido del papel rompiéndose. Como resultado, las neuronas espejo audiovisuales, que responden a sonidos de acciones específicas, se activaron. Así, este sistema permite la comprensión de las acciones independientemente de si estas son vistas, escuchadas o realizadas.

La comprensión sobre cómo los individuos pueden tener información de las intenciones de los otros se encuentra en los circuitos neuronales. Las neuronas que codifican actos motores individuales —como sujetar— se encuentran insertas en circuitos previamente formados, cada uno de los cuales codifica una acción específica (las acciones son secuencias de actos motores individuales). De esta manera, las neuronas se vinculan con otras que codifican los próximos actos motores y, posiblemente, los facilita. Es el contexto el que da las claves sobre qué acción es la que se está realizando. En un experimento se observó que las neuronas espejo del lóbulo parietal inferior se activaban de manera diferenciada cuando el acto motor de agarrar pertenecía a distintas acciones —agarrar para comer y agarrar para colocar—. Las neuronas espejo, entonces, permitieron predecir el siguiente acto motor; esto es, la intención (2007).

La posibilidad de captar la intención de las demás personas es un aspecto fundamental para la interacción social. Si bien aún queda más por descubrir con respecto a las neuronas espejo, se sabe ya, por ejemplo, que el asco se encuentra dentro de este sistema. Es posible que diversas emociones también estén codificadas en el sistema de neuronas espejo, lo cual brindaría una comprensión más cabal de la empatía y daría cuenta de ciertas bases fundamentales para las relaciones interpersonales.

1.3. La atribución psicológica en el desarrollo social primate

Como vimos, desde sus orígenes la teoría de la selección natural ha servido, entre otras cosas, como paradigma para la investigación de la historia biológica de nuestras capacidades cognitivas y sociales. Ello implica que algunos aspectos de aquellas ya se encontraban, de forma quizás más elemental, en especies anteriores. Esto motiva a la primatología a producir y contrastar hipótesis sobre posibles capacidades compartidas con los grandes simios, pues se asume que, de las especies que viven en el presente, son las más cercanas a nosotros y, por tanto a nuestros antepasados. De hecho, uno de los textos que acuñaron el término *teoría de la mente* es el ya clásico artículo «Does the Chimpanzee Have a Theory of Mind?» de David Premack y Guy Woodruff (publicado por primera vez en 1978). Mediante dicho término se hace referencia a la capacidad para atribuirse a uno mismo estados mentales y de, además, atribuirles a otros individuos estados mentales que podrían ser distintos a los de uno.

Posteriormente, en «Does the Chimpanzee Have a Theory of Mind? Revisited», Premack (1988) evalúa los métodos y los resultados de las investigaciones en torno a si aquellos primates superiores pueden, o no, realizar atribuciones mentales de manera similar a los humanos. El desarrollo de estudios semejantes implica resolver o, al menos, tomar partido respecto de dos cuestiones más básicas. La primera se refiere al rol del lenguaje en el desarrollo de la mente. Hoy en día, muchos autores proponen formas de lidiar con este problema. Por ejemplo, Peter K. Smith piensa que el único modo seguro de identificar la atribución de estados mentales depende eminentemente de un sistema simbólico como el lenguaje humano. Jim Hopkins sostiene que el surgimiento de las teorías de la mente se explica a partir de la forma en la que empleamos el lenguaje para describir el entorno, al punto que nuestra forma de hablar de los estados mentales es muchas veces análoga a la que empleamos para hablar de eventos físicos. Del otro lado del espectro, para Peter Carruthers la explicación tanto evolutiva como operativamente más económica de la atribución mental debe implicar la idea de que los mecanismos metarrepresentacionales (la capacidad de representarse las representaciones ajenas) pueden operar en instancias previas al empleo del lenguaje. Por su parte, Premack, como se verá con más claridad en lo que sigue, considera que es necesario distinguir entre formas de atribución, y algunas de ellas no dependen del lenguaje.

La segunda cuestión ocupa el centro de la atención de Premack a lo largo del escrito. Se trata del valor probatorio de los experimentos en circunstancias controladas, frente al valor de las observaciones hechas en hábitats naturales (1988, p. 161). Cabe preguntarnos si los experimentos acaso no generan circunstancias tan poco naturales que las respuestas de los individuos así estudiados corren el riesgo de no explicar nada que ya haya sido puesto previamente por quienes indagan.

Un tipo de caso registrado que suele vincularse con la tesis de que los chimpancés podrían poseer alguna especie de teoría de la mente es el engaño. Lo interesante del caso radica en que estos primates son capaces de engañar en una gama muy amplia de circunstancias. El ejemplo de engaño que emplea Premack es el de la omisión del gemido que los chimpancés macho suelen emitir cuando copulan con su hembra preferida. Ello tiene como objeto no llamar la atención de otros machos para no tener que compartir la hembra. Ahora bien, el problema consiste en determinar si el individuo que emplea dicha estrategia lo hace teniendo en cuenta que con ello evita que los demás *sepan* algo y, por tanto, evita que actúen de cierta manera no deseada, o si meramente busca evitar que los demás *hagan* algo. Solo el primer escenario instancia necesariamente una teoría de la mente por parte del chimpancé, mientras que en el segundo bastaría con una explicación conductista. El autor niega que para poder discernir entre ambas posibilidades ayude recurrir a información obtenida en otras experiencias de campo, pues la nueva información también se podría interpretar de dos maneras distintas. Aquí es donde la experimentación controlada resulta ser particularmente importante. En opinión de Premack (1988, p. 171) (y, como diremos más adelante, de Whiten también) únicamente aquella nos permite realizar distinciones conceptuales finas. Frente al contraargumento del condicionamiento que podría tener lugar en un laboratorio, factor que podría volver la prueba demasiado forzada, este responde que tampoco en el campo existen garantías que protejan a los individuos contra ello.

Gracias a la experimentación en psicología del desarrollo, se sabe que los niños de tres años ya son capaces de atribuir estados mentales a otros agentes y, sin embargo, no comprenden aún la relación entre sensaciones y creencias, en particular cuando se trata de los demás (Premack, 1988, p. 164). Al parecer, esta capacidad se adquiere cabalmente a partir de los cuatro años de edad, cuando ya se posee una teoría de la mente avanzada. Ya que los chimpancés fallan pruebas sobre el papel de las sensaciones en la adquisición de conocimiento que niños de tres años y medio tampoco aprueban, Premack piensa que se puede postular como regla general que si un niño de dicha edad no puede hacerlo, entonces un chimpancé tampoco. No obstante, ciertos experimentos parecen demostrar que algunos chimpancés sí son capaces de predecir el comportamiento de otros agentes o emplear tácticas en su beneficio basándose en lo que podría ser alguna forma de conexión del comportamiento y las sensaciones. Para explicar estos contraejemplos, piensa el autor, debemos afinar la idea de atribución de estados mentales.

Según Premack, resulta necesario distinguir entre dos tipos de estados mentales, unos más complejos y otros menos. Los estados más simples son automáticos, encapsulados y poco variables. Las sensaciones se encontrarían dentro de este grupo.

La creencia, de otra parte, parece ser del tipo complejo, pues no solo es altamente variable, no automática, ni encapsulada, sino que, además, se representa mediante el lenguaje. La creencia también permite poner en duda los estados sensoriales que de otra manera estarían ligados directamente a la acción. Cuando vemos a un individuo humano vacilar, pensamos inmediatamente que, por ejemplo, se está preguntando si realmente ve lo que cree ver. Se tiene registro de vacilación en chimpancés, pero ello podría tratarse del resultado de un condicionamiento mixto, es decir, de que los primates hayan estado expuestos a experiencias diversas con resultados unas veces positivos y otras veces negativos respecto de la misma cosa. En general, las probabilidades de encontrar una explicación conductista que reemplace una postura mentalista respecto de la tenencia de las creencias simples son bastante altas.

También poseemos una gran cantidad de creencias que admitimos como posibles explicaciones sin que necesitemos recurrir, al menos directamente, a datos sensoriales. Premack (1988, p. 174) cita ejemplos como la idea de que la mente está en el cerebro, los gérmenes causan enfermedades o que las especies evolucionan. Estas creencias parecen depender de dos factores: una competencia lingüística considerable y la capacidad de realizar explicaciones causales de manera consciente. En estos casos, la brecha entre humanos y el resto de animales parece infranqueable. No obstante, el autor considera que este es un campo abierto a la investigación. Un ámbito particular donde se puede profundizar en esto es en la manera como humanos y otros primates generan (o no) creencias y expectativa respecto de la conducta de los demás. En los primeros, un encuentro con otro individuo puede producir predicciones de dos tipos: de una parte, predicciones basadas en el mero condicionamiento, de manera que frente a un rango restringido de casos se tiene la *expectación* de que el otro haga algo; de otra, los humanos también, a raíz del encuentro, adoptamos *creencias* acerca de la personalidad del otro en general, lo cual nos permite realizar predicciones en un rango bastante más amplio de contextos posibles. La experimentación ha mostrado que los chimpancés jóvenes pueden llegar a hacer predicciones basadas en la expectativa, mas parecen carecer de juicios más generales, aquellos del segundo tipo.

Hasta aquí, pareciera que los chimpancés solamente son capaces de atribuir estados mentales simples, como la expectativa o los otros estados sensoriales. De hecho, algunos experimentos muestran que estos primates no son capaces de distinguir entre sus propios estados mentales, como lo que ellos ven, y los estados de otros individuos. Esto contrasta notoriamente con la facilidad con que los niños de cuatro años aprueban muchas versiones del test de la falsa creencia, el cual es la principal herramienta para determinar la tenencia de teoría de la mente en humanos. La idea básica, desarrollada por Wimmer y Perner (1983), consiste en colocar a los infantes frente a una situación en la que deberían atribuirle a otro agente creencias basadas

en una percepción que ya no corresponde al estado de cosas actual. Por ejemplo, en la primera versión de la prueba (*The Maxi Task*) se les contaba un relato a los niños, en el cual el personaje principal, Maxi, colocaba su barra de chocolate en un lugar específico y salía a jugar. Luego, su madre cambiaba el chocolate de lugar y Maxi regresaba a buscarlo. Se les preguntaba a los niños dónde creían que Maxi buscaría su chocolate, si acaso lo haría donde originalmente lo dejó o en el nuevo lugar al que su madre lo movió. Antes de los cuatro años, los niños suelen errar la respuesta al afirmar que Maxi buscará el chocolate donde ellos saben que está (en el nuevo lugar asignado por la madre), en vez de responder poniéndose en el lugar del otro agente. No obstante, como hemos visto, la cuestión por la atribución psicológica en primates no humanos sigue abierta, y el refinamiento en los métodos de experimentación controlada resultará decisivo para el progreso de la investigación.

A pesar de lo dicho arriba respecto de las dudas en torno de la tenencia plena de teoría de la mente por parte de primates no humanos, Robin Dunbar (1988) piensa que se les puede atribuir intencionalidad de segundo orden, lo cual quiere decir que cuentan con estados mentales de la forma «yo *creo* que tú *crees* que...», en la cual el verbo mental («creer», en este ejemplo) es asignado en dos niveles distintos («yo» y «tú»). No obstante, los humanos podemos llegar considerablemente más lejos en el conteo de asignación de estados mentales. Si bien no existe consenso en torno al número máximo de niveles a los que la intencionalidad humana puede llegar, las pruebas indican que entre el cuarto y el quinto nivel de asignación de creencias la tasa de errores de los individuos experimenta un crecimiento dramático de entre 5% y 10% para el cuarto nivel a alrededor de 60% para el quinto nivel.

Según Dunbar, deberíamos estar en condiciones de elaborar un relato evolutivo que dé cuenta de la diferencia al menos cuantitativa entre los niveles de intencionalidad humanos y los de otros primates. La diferencia, además, parece ser compatible con algunos datos obtenidos de la comparación de la constitución del cerebro en monos y homínidos. Aparentemente, el tamaño de la sección del cerebro de los monos dedicada a procesar información visual, «V1», crece en paralelo con las demás áreas del neocórtex, dedicadas al procesamiento de, entre otras cosas, modos reactivos del comportamiento. Sin embargo, la proporción cambia en el caso de los homínidos, en los que las otras áreas del neocórtex se desarrollan bastante más que V1. Esto, a su vez, tiene un correlato en el comportamiento. Mientras los monos pueden leer y predecir el comportamiento de otros individuos a partir del procesamiento de *inputs* visuales, no son capaces de «leer la mente detrás del comportamiento» (1988, p. 245), cosa que los homínidos sí parecen poder (ya vimos que el ser humano en grados más elevados que el resto de las especies emparentadas).

La habilidad de lectura de mentes de los homínidos, entonces, parece depender del tamaño del neocórtex. La complejidad de un cerebro apto para producir teorías

de la mente refinadas es bastante alta, lo cual trae consigo costos biológicos elevados para mantener semejante órgano en funcionamiento. Por tanto, desde un punto de vista evolutivo, estas características cerebrales deberían haber sido seleccionadas directamente debido al gran beneficio que traen consigo para sus especies portadoras. Dunbar y muchos otros ven una correlación entre el desarrollo del cerebro en los homínidos y el número de individuos que componen sus respectivos grupos sociales. Como ya debe sospecharse, dada la función social de la lectura de mentes, es este dispositivo y otros semejantes los que se piensa que contribuyen a la organización de sociedades numerosas.

Casi todos los elementos del relato sobre la evolución de la capacidad para la teoría de la mente están listos; solo hace falta identificar la presión selectiva que impulsó el desarrollo de sociedades de homínidos cada vez más numerosas. Dunbar (1998) piensa que fue probablemente la necesidad de encontrar y mantener acceso al agua lo que ejerció la presión necesaria para la evolución de habilidades sociales más complejas. Esta presión habría ido en aumento desde que, con la aparición del *Homo erectus* (hace aproximadamente 1,5 millones de años), nuestros antepasados homínidos empezaron a desplazarse grandes distancias. Entonces habría sucedido un incremento en el tamaño del neocórtex y un correspondiente incremento en el número de integrantes de los grupos, excediendo el límite superior de las sociedades actuales de grandes simios. De la misma manera, se estima que es con la aparición de las primeras formas de *Homo sapiens* (hace aproximadamente 120 000 años) que se dio otro salto, tanto respecto de la constitución del cerebro como en la cantidad de integrantes de cada grupo.

1.4. La hipótesis de la inteligencia social

Naturalmente, poder predecir las acciones de los demás con vistas a obtener el mayor beneficio posible para uno, resulta de gran utilidad adaptativa. Esta capacidad, la *inteligencia maquiavélica*, se ha desarrollado entre los humanos gracias a la manera en que usualmente predecimos el comportamiento de los demás, es decir, gracias a que atribuimos estados mentales a otros agentes. Por ello, para poder determinar si otros primates poseen también inteligencia maquiavélica, debemos responder primero si es que son capaces de leer mentes, al menos en un nivel elemental.

1.4.1. Inteligencia maquiavélica y niveles de atribución psicológica

Aparentemente, las pruebas de la falsa creencia sirven como el mejor criterio para determinar si el individuo ya posee mentalismo considerablemente desarrollado o no. Como vimos, Premack (1988) sostiene que los experimentos en chimpancés solo alcanzan para mostrar que, dada una secuencia de acciones de un agente, los primates

pueden predecir correctamente la acción siguiente, pero porque es lo que ellos mismos harían. El mentalismo que buscamos requiere, en cambio, que el individuo sea capaz de adscribirle a otro agente estados mentales distintos de los suyos. A pesar de estos resultados, Andrew Whiten (1997) piensa que estudios realizados acerca de la atribución de ciertas clases de estados mentales en niños de edades anteriores a la requerida para aprobar las test de la falsa creencia, resultan de particular interés en la investigación del maquiavelismo de primates no humanos. Cabe recordar lo que dijimos que Premack (1988, p. 146) consideraba una regla no estricta pero útil: si el niño de tres años y medio no puede, tampoco puede el chimpancé. Pero, en el caso contrario, si el niño pequeño puede, entonces podría ocurrir que otros primates también.

Ya que se puede rastrear formas de atribución mental previas al mentalismo avanzado requerido para aprobar las pruebas de la falsa creencia, Whiten (1997, p. 150) sostiene que es preferible pensar la lectura de mentes gradualmente. Así, distingue entre cuatro posibles tipos. La forma más elemental sería aquello que llama *lectura implícita*, y se refiere a la capacidad de identificar un nexos entre la percepción de un agente y la acción, relación mediada, no explícitamente, por una mente. Un segundo tipo es el *contraengaño*, que consiste en la habilidad para reconocer las intenciones detrás de las acciones realizadas en falso por un agente. El *reconocimiento de variables intermedias* implica saber que ciertos comportamientos generan estados mentales en los demás agentes, lo cual, a su vez, permite realizar predicciones. Finalmente, la *proyección de la experiencia*, como el nombre sugiere, consiste en saber utilizar la propia experiencia en la predicción del comportamiento de otros. Whiten mismo cree que los últimos dos tipos de mecanismos son indicios más seguros de mentalismo, aunque, como dijimos, las otras opciones no pueden ser descartadas sin más.

Ya los niños de casi tres años emplean mayoritariamente explicaciones causales mentales para describir sus relaciones con el entorno. Y no solo ello, sino que dichas explicaciones son empleadas, más que en cualquier otra circunstancia, para extraer beneficios personales. De hecho, se ha registrado que son los niños que más hábilmente emplean su capacidad de lectura de mentes en contextos maquiavélicos los que posteriormente poseerán una teoría de la mente más desarrollada (p. 153).

El maquiavelismo precoz de los niños parece contradecir los resultados de las pruebas de la falsa creencia. Además, experimentos controlados demuestran que los niños pequeños sí son engañadores capaces. El problema de lo último radica en que también se sabe, mediante experimentación, que los niños de tres años, a pesar de su capacidad para ciertas formas de engaño, no extraen todas las consecuencias mentalistas que se esperaría de alguien que comprende cabalmente en qué consiste poseer una representación falsa (como la que ellos mismos logran crear con sus tácticas engañadoras). Al parecer, la razón de ello radica en que los pequeños de tres

años aún no poseen un concepto completo de creencia, y sus estrategias de engaño, en el fondo, son intentos por hacer que el agente a quien quieren aventajar participe en una realidad pretendida, para lo cual no es necesario que comprendan en qué consiste creer falsamente (p. 156). Así, los contextos maquiavélicos tempranos sirven para desarrollar y afianzar mecanismos necesarios para el surgimiento del mentalismo maduro. Tal vez aquí es donde puede encontrarse un paralelo con el maquiavelismo de primates no humanos.

No se tiene evidencia concluyente sobre hasta qué punto los niños pequeños y los primates no humanos distinguen entre la realidad percibida y los estados mentales de los demás. En los niños, el problema se resuelve a partir de los cuatro años. No obstante, otros primates también exhiben comportamiento engañoso. Al igual que Premack, Whiten se pregunta si dicho comportamiento no se puede explicar recurriendo al mero conductismo. Mas en torno a este punto, nuevamente se puede establecer un paralelo entre lo encontrado en psicología del desarrollo con relación a los niños y algunos datos experimentales de la primatología. Nos referimos específicamente a la capacidad de leer la atención que otros agentes le prestan al entorno compartido. Si bien la información que se tiene acerca de ello en chimpancés y otros primates no es consistente, Whiten enfatiza el hecho de que en ciertas pruebas los chimpancés intentan trasladarse físicamente a un punto donde puedan visualizar qué cosa está observando otro agente. Esto podría indicar que, de hecho, los individuos así estudiados saben que la visión es siempre *acerca de algo*. La intencionalidad de la visión es considerada, además, como una posible primera puerta de los infantes hacia la teoría de la mente, pues es esencial en su aprendizaje sobre cómo nombrar objetos mediante la atención que le prestan al mismo acto en sus cuidadores (pp. 164-165). Por motivos similares, el autor también considera que resulta de igual importancia para el estudio de la inteligencia maquiavélica determinar en qué medida los primates no humanos pueden distinguir entre posibles voliciones de otros agentes.

Las conclusiones a las que puede llegar Whiten luego de considerar toda la información mencionada, son prácticamente las mismas a las que llega Premack. La investigación sobre la atribución mental en primates no humanos se enriquece mucho con lo que la psicología del desarrollo descubre en los infantes, pero aún quedan muchos cabos sueltos por resolver. Esta tarea se debe realizar, sobre todo, con la ayuda de la experimentación controlada. Sin embargo, a pesar del estado aún embrionario de esta empresa científica, Whiten también considera que, en un sentido mucho más simple que en el relativo a los humanos, se aprecia en otros primates alguna forma de atribución mental. O, al menos, se aprecian mecanismos necesarios para el desarrollo posterior de teorías de la mente.

1.4.2. Complementos a la hipótesis de la inteligencia social

Si bien la hipótesis de la inteligencia social ha abierto un amplio campo de investigación evolutiva, podemos preguntarnos sobre sus límites explicativos (Byrne, 1997). Podríamos pensar que dicha idea basta para explicar las diferencias en la evolución de las capacidades cognitivas de ciertos primates, incluidos los humanos. Pero también puede pensarse que la teoría no es suficiente y, entonces, sea necesario indagar otras causas complementarias del fenómeno evolutivo.

Como vimos al discutir algunas tesis de Dunbar, parece haber una correlación entre el incremento del tamaño del *neocórtex* y la complejidad de la vida social de las especies. Naturalmente, la hipótesis de la inteligencia maquiavélica se ve reforzada por esta evidencia. Específicamente, datos de este tipo se ajustan muy bien en el caso de las diferencias entre los primates haplorrinos, como los grandes simios o los monos, y los estrepsirrininos, como los lémures. No obstante, la relación entre los datos y la teoría no siempre es armoniosa. Al interior de la rama de los haplorrinos hay diferencias importantes: existen diferencias notorias entre las capacidades cognitivas de simios y monos, pero las sociedades de los primeros no son necesariamente más complejas ni más grandes que las de los segundos. De hecho, un aspecto importante de la postura de Dunbar era la proporción del tamaño del neocórtex con relación al resto del cerebro, mas dicha proporción no varía mucho entre los haplorrinos (Byrne, 1997, pp. 294-295).

Byrne sostiene que las principales diferencias cognitivas entre los grandes simios y otras especies del mismo suborden no se manifiestan únicamente en las habilidades sociales. Sin embargo, sí resulta relevante notar que, por ejemplo, el engaño, estrategia tan importante para el estudio de la inteligencia maquiavélica, aunque presente también en monos, es ejecutado de manera cualitativamente distinta por los grandes simios. En estos parece haber cierto grado de planificación que, como también vimos en secciones anteriores, podría indicar la presencia de alguna forma débil de atribución psicológica.

De manera similar al engaño y otras estrategias sociales, si bien tanto monos como simios superiores son capaces de utilizar ciertas herramientas físicas elementales, los primeros no parecen ser capaces de representarse específicamente qué características del objeto seleccionado son relevantes para aquella actividad (usualmente extracción de material alimenticio) en la que lo emplean. Los segundos, en cambio, aparentemente sí son capaces de hacer este tipo de distinciones finas desde el momento mismo de la selección de los objetos (1997, p. 296). Otra habilidad que llama la atención de los estudiosos, y que se relaciona con el uso de herramientas, es la construcción de camas o nidos por parte de gorilas (de manera más simple), y orangutanes y chimpancés (con mayor complejidad). Curiosamente, esta habilidad está presente en otros primates, pero no en haplorrinos.

Byrne también menciona la capacidad de imitar como factor diferenciador entre los grandes simios y otros primates (p. 297). Incluso dándole una interpretación modesta (que no recurre a la postulación de tenencia de alguna forma de teoría de la mente), la imitación parece requerir de la capacidad de representación del comportamiento en general, lo cual solo parece presente, nuevamente, en los primeros.

Todo indica que aquello que distingue a los grandes simios es una cierta capacidad de abstracción. Sus respuestas más finas a ciertos problemas parecen implicar, al menos, que los individuos estudiados pueden simular escenarios posibles. Estas facultades, además, no se limitan a la esfera de los problemas sociales, esfera en la que los monos también son bastante hábiles. Surge, entonces, la pregunta por las presiones selectivas que distanciaron cognitivamente a los grandes simios del resto de primates.

Algunas posibles explicaciones de la evolución peculiar de los grandes simios se centran en sus limitaciones motrices en tanto braquiadores, si se tiene en cuenta que los otros primates son considerablemente más ligeros. Así, la selección natural habría favorecido en aquellos una mayor capacidad de procesamiento de datos espacio-temporales o, en otras palabras, un mapa cognitivo más complejo (pp. 298-299). Arriba hablamos acerca de la capacidad de simulación, la cual resulta particularmente útil considerando que los grandes simios, dado su peso, corren bastante más riesgo de caer que los otros braquiadores arbóreos más ligeros. Naturalmente, la simulación permitiría al individuo ponerse en casos posibles y anticipar sus propios movimientos, reduciendo el riesgo de caída. Por ello, la simulación del movimiento propio es otra posible capacidad que determinó la evolución cognitiva de los grandes simios (p. 300). De otra parte, su tamaño también podría haber influenciado en la evolución de sus capacidades representacionales, imponiendo la necesidad, ausente en el resto de primates, de desarrollar técnicas sofisticadas de construcción de camas en los árboles (posiblemente en un antepasado común a los grandes simios contemporáneos). Esta inteligencia técnica básica luego podría haberse desarrollado en otros ámbitos de la vida de dichos simios, como se observa hoy.

Se han desarrollado teorías en torno a la sorprendente capacidad de los grandes simios para emplear herramientas que los ayudan a extraer alimentos. Dichas teorías, sin embargo, se topan con el problema de que el uso de herramientas también está presente en algunas especies de monos. Byrne piensa que los problemas se disipan si nos enfocamos, más allá del mero hecho del empleo de herramientas, en la manera en la que los individuos lo hacen (p. 303). De hecho, si el empleo de herramientas en los grandes simios responde a esquemas sistemáticamente organizados de acción, entonces cabría esperar esquemas semejantes en actividades en las que se prescindiera de herramientas. Al parecer, sí puede observarse comportamiento organizado de este tipo en gorilas y orangutanes, específicamente en sus técnicas manuales para la obtención de alimento.

En el caso de los monos, sus técnicas manuales para la obtención de alimento no parecen responder a una organización sistemática. Por el contrario, estas parecen ser producto de una rápida capacidad de aprendizaje volcada sobre ciertos estímulos repetidos. Los grandes simios, en cambio, presentan secuencias de acciones organizadas jerárquicamente de acuerdo con objetivos primarios y subordinados. Este refinamiento en las técnicas de obtención de alimento parece denotar la capacidad de representación del comportamiento, y se debería a las exigencias alimenticias que mamíferos grandes con estómagos no particularmente resistentes deben satisfacer en hábitats muchas veces complejos.

La propuesta tentativa de Byrne sería, sintéticamente, que algún antepasado común de los grandes simios (probablemente similar a los gorilas contemporáneos) desarrolló la capacidad de organizar distintos aspectos de su comportamiento de acuerdo con planes bastante flexibles que le permitan superar limitaciones físicas no presentes en otros primates. Esto último, evidentemente, implica cierto nivel de entendimiento representacional. Ello podría ayudar a explicar, además, el incremento en el tamaño absoluto del cerebro de los grandes simios, ya que las exigencias computacionales de un nivel semejante de planificación son bastante más costosas que en el comportamiento de, por ejemplo, los monos.

1.5. Evolución de la lectura de mentes, la autoconciencia y la agencia

En años recientes, las teorías que explican la evolución de la cognición social también han sido empleadas como nuevas formas de enfrentar problemas clásicos de disciplinas como la filosofía y la psicología. En las siguientes secciones, trataremos dos casos interesantes: el de la explicación filogenética de la relación que existe entre el conocimiento de nuestras propias mentes y de las mentes de los demás, y el problema de explicar nuestra experiencia como seres capaces de modificar su propio comportamiento voluntariamente.

1.5.1. La relación de la lectura de mentes y la metacognición. Explicaciones filogenéticas

Por lo general, en circunstancias normales, los seres humanos pensamos que poseemos la facultad de acceder directamente a nuestros propios estados mentales. Aquí, *acceder directamente* quiere decir que el conocimiento que poseemos sobre lo que ocurre en la mente de uno mismo es considerablemente menos falible que aquello que conocemos perceptualmente. Por ejemplo, los estados mentales que creemos leer en el comportamiento de otros agentes, en relación con los cuales, por ello mismo, tan solo seríamos intérpretes. Pero, si pensamos que es gracias al mismo dispositivo, alimentado por fuentes similares de información, que interpretamos a los demás

y conocemos nuestras propias mentes, la asimetría entre la falibilidad de ambas actividades se disipa. Esta última es la tesis de Carruthers (2009), quien considera que tanto la metacognición como la lectura de mentes son actividades del mismo dispositivo metarrepresentacional.

A la idea del acceso directo a las propias representaciones se le suele llamar *introspección*, y es este mismo sentido del término el que Carruthers intenta abandonar. Según su modelo, todo lo que se necesita para explicar la metacognición es que el dispositivo metarrepresentacional, cuya función primaria es la lectura de mentes, tenga acceso a los datos de salida (*outputs*) de todos los otros dispositivos cognitivos que procesan la información perceptual y cuasi perceptual (como los productos de la imaginación visual o el discurso interno) de primer orden. Esto quiere decir que el dispositivo de lectura de mentes interpreta el comportamiento de los agentes sobre la base de los conceptos de primer orden que los otros dispositivos han utilizado. No obstante, jamás se tiene información perceptual sobre el proceso de generación de juicios. Vemos, por ejemplo, a un agente caminar, y el dispositivo de lectura de mentes recibe la información perceptual ya procesada con conceptos como «persona», «caminar» y todos los que sean necesarios, pero no recibe ningún dato sobre las creencias que impulsan la acción del caminante. Exactamente lo mismo ocurre en el caso de la primera persona: tampoco poseemos acceso consciente a los procesos de generación de nuestros propios juicios y decisiones. Justamente, este es el espacio que le corresponde completar al dispositivo metarrepresentacional.

El caso de la interpretación de uno es, prácticamente, el mismo. La única diferencia, cuando la hay, consiste en que la mayor parte del peso de los datos perceptuales que el dispositivo metarrepresentacional debe interpretar recae en aquello que llamamos más arriba *cuasi perceptual*. De esta manera, puede que la atención que le prestemos al entorno en un momento determinado sea mínima, pero seguimos produciendo expresiones verbales internas e imágenes visuales basadas en recuerdos. Esta información cuasi perceptual será la que sirva como principal punto de apoyo para la labor del dispositivo metarrepresentacional, con lo que surge la metacognición. Así también se explica la idea de la introspección. Es este enfocarnos en los datos cuasi perceptuales durante la metacognición lo que nos hace pensar que ello depende de un mecanismo esencialmente distinto a cualquier otro que emplee información del entorno.

De todo lo anterior se siguen consecuencias evidentes. La metacognición resulta ser tan interpretativa como la lectura de mentes, pues se trata de dos caras de un mismo mecanismo. Por lo mismo, ya no es necesario postular un dispositivo especial de introspección que garantice la metacognición; aquella actividad es, en el fondo, uno de los conceptos que conforman la teoría *folk* de nuestras mentes, pero no es una función real. Se predice, además, que somos altamente confabuladores con nosotros mismos.

Esto quiere decir que, en aras de un funcionamiento más eficiente, el dispositivo metarrepresentacional puede interpretar en falso el comportamiento de uno mismo (de hecho, acabamos de afirmar que la idea misma de un mecanismo encargado de la introspección es falsa, aunque computacionalmente provechosa).

Como vimos en la introducción a este capítulo, Carruthers piensa que la refutación final de las teorías que dependen de la existencia de un mecanismo introspectivo, como las teorías de la simulación, debe realizarse en el terreno de las explicaciones filogenéticas de los mecanismos sobre los que se discute. La razón es que, aparentemente, la información proveniente de los estudios ontogenéticos de nuestros mecanismos para la interpretación del comportamiento de los agentes puede ser interpretada desde más de un modelo, ya sea uno afín a las TT o a las TS.

El argumento filogenético a favor del modelo que postula un solo mecanismo metarrepresentacional para la metacognición y la lectura de mentes es sumamente sencillo. La idea consiste en que, desde la perspectiva de la teoría de la evolución, resulta mucho más económico postular una sola historia evolutiva para más de una función cognitiva. Biológicamente, la generación y el mantenimiento de los mecanismos implicados en la atribución psicológica son sumamente costosos. Si se postula un mecanismo metacognitivo (con capacidades introspectivas) aparte de uno de lectura de mentes, es decir, de atribución en tercera persona, se deben postular dos historias evolutivas distintas para cada mecanismo, cada uno biológicamente muy costoso. Ahora bien, otra opción, entre tantas, es sostener que solamente se requiere de un tipo de mecanismo que se encargue de todas las formas de atribución psicológica, pero que este mecanismo fue originalmente diseñado para la metacognición. Esta es la postura de Alvin Goldman (2006), defensor de las TS. Contra Goldman, Carruthers argumenta que una tesis semejante depende también de al menos dos historias evolutivas diferentes. De una parte, se tendría que dar cuenta de qué presiones evolutivas actuaron en la selección de un mecanismo metacognitivo anterior e independientemente de la metarrepresentación; de otra, se tendría que ofrecer una explicación de bajo qué presiones evolutivas aparecieron mecanismos que permitan identificar la información perceptual de otros agentes con las propias representaciones metacognitivas.

Según Carruthers, el modelo de un solo mecanismo metarrepresentacional es el que mejor compatibiliza con la hipótesis de la inteligencia maquiavélica, tal y como es postulada por autores como Premack o Dunbar. La complejidad de la vida social de nuestros antepasados homínidos parece ser la mejor alternativa de presión evolutiva para explicar la selección de nuestras capacidades metarrepresentacionales. Por ello, la historia evolutiva que imagina Carruthers parte de la selección de un mecanismo encargado de la lectura de mentes. Es decir, de un mecanismo metarrepresentacional *seleccionado para* la lectura de mentes. La metacognición sería,

entonces, un producto colateral de la evolución del dispositivo de lectura de mentes. De esta manera, obtenemos una misma historia evolutiva para explicar dos facultades cognitivas diferentes.

1.5.2. El problema de la causalidad mental

Propuestas como la de Carruthers no solo ponen en tela de juicio conceptos como el de introspección. También, ponen en duda otras intuiciones típicas de nuestra *psicología folk*, ciertamente vinculadas con el concepto de introspección, como la idea de que muchas veces nuestro comportamiento es causado por nuestra voluntad luego de un proceso deliberativo. Normalmente, creemos que tomamos decisiones y actuamos acorde a ellas. Pero, si la vía de acceso al conocimiento de la primera persona es casi tan falible como la del conocimiento de la tercera persona, si constantemente confabulamos contra nosotros mismos, entonces podemos dudar de nuestra experiencia cotidiana de la voluntad como causa de muchos actos.

Lo dicho arriba es una expresión contemporánea de un clásico problema filosófico: el de si es posible justificar alguna forma de compatibilidad entre la causalidad de los fenómenos naturales, tal como los describimos en las ciencias, con la idea común de que podemos ejercer un control consciente sobre nuestros actos. Quienes defienden que los eventos mentales, como las creencias y los deseos, sí se pueden pensar como causas legítimas de eventos físicos, deben demostrar que los nexos causales descritos en términos mentales no son reductibles a descripciones estrictamente físicas. Este tipo de defensas van de la mano de explicaciones sobre la eficacia de las descripciones mentales del comportamiento. En lo que sigue, veremos dos ejemplos recientes de dicho modo de defender la causalidad mental.

Si bien el fisicalismo (la idea de que toda propiedad y evento del universo son, al menos en una de sus posibles descripciones, propiedades y eventos físicos) es la tesis ontológica más aceptada en la filosofía contemporánea, pareciera que si se le lleva a sus últimas consecuencias nos obliga a renunciar a otras tesis que suelen considerarse importantes para explicar ámbitos específicos de la realidad. En particular, pareciera obligarnos a aceptar que todo sistema no descrito en términos de las leyes de la física es reducible, en principio, a esta. Una de las formas de señalar el problema es mediante la discusión de la idea de *superveniencia* de propiedades no físicas a propiedades físicas y su efectividad o irrelevancia causal. En filosofía, el término *superveniencia* se utiliza para hablar de la dependencia de ciertas propiedades respecto de otras más básicas. Así, por ejemplo, si admitimos, como sería consecuente con el fisicalismo, que las propiedades psicológicas de un agente *supervienen* a propiedades físicas, cualquier cambio en las primeras implica un cambio correspondiente en las segundas. Si aceptamos la superveniencia de lo psicológico sobre lo físico, entonces al considerar

que una propiedad psicológica cualquiera, digamos p , causa la aparición de otra propiedad psicológica p' ; y aceptamos que p superviene a una propiedad física f y p' a una propiedad física f' ; entonces la descripción del nexo causal entre p y p' puede reducirse a la descripción del nexo entre f y f' . En términos menos formales, al reduccionismo fiscalista subyace la idea de que, puesto que toda propiedad no física superviene a alguna propiedad física, toda descripción no física es, en principio, reducible a otra física y las relaciones causales no son una excepción a la regla.

Como notan El-Hani y Pereira (1999), el fiscalismo reduccionista entra en conflicto con una creencia cuya verdad resulta difícil de cuestionar: que los sistemas que explican distintos niveles de la realidad, como aquellos que componen las teorías de la biología o de la psicología, deben su relevancia a que responden a un dominio propio de cuestiones, por lo que debería garantizarse su independencia causal respecto de las propiedades meramente físicas. Por ello, los autores mencionados se proponen demostrar que el concepto de *emergencia* de niveles de organización más compleja de propiedades a partir de niveles menos complejos resulta útil para solventar un fiscalismo no reduccionista.

Los sistemas biológicos, por ejemplo, emergen de sistemas físicos y químicos menos complejos. En un sentido, existe continuidad entre los niveles, pues en todos casos se trata de los mismos componentes materiales. Las partículas que estudia la física son las mismas que conforman los organismos que estudia la biología. De hecho, el concepto de emergencia no puede desligarse de la idea de que los distintos niveles de organización han ido surgiendo en una misma historia evolutiva. No obstante, a pesar de poseer el mismo sustrato material, entre los distintos niveles existe una discontinuidad de propiedades, correspondiente a sus diferencias en la complejidad de organización. De ahí que el concepto de emergencia sea compatible con el de *dualismo de propiedades* (o, quizá, sea más pertinente hablar de un *pluralismo* de propiedades) (El-Hani & Pereira, 1999, p. 337).

El-Hani y Pereira sostienen que el problema de la irrelevancia causal de las propiedades supervenientes se mantiene si es que solo se admite la idea de causalidad eficiente (1999, p. 340). Piensan que el problema puede ser afrontado si se toma en cuenta una reinterpretación contemporánea de la concepción aristotélica sobre los cuatro tipos de causalidad. De esta manera, se piensa la causalidad funcional, correspondiente a la causalidad final aristotélica, como referida a la función que un elemento juega dentro de su correspondiente sistema con relación a la mantención de su estabilidad. Asimismo, la causalidad formal aristotélica se reinterpreta en términos de los patrones de coordinación que los elementos de un sistema adoptan. La causalidad material se refiere, ahora, a las propiedades de dichos elementos, propiedades que, hemos visto, supervienen sobre las de elementos de un nivel inferior.

Lo anterior nos debería permitir evitar comprender unidireccionalmente las relaciones causales entre los distintos niveles organizativos (unidireccionalmente, pues los nexos causales de los niveles superiores son, de acuerdo con el fisicalismo reduccionista, simples epifenómenos). La posibilidad de la reducción es la otra cara de una forma de esencialismo, esto es, de la creencia de que, en el fondo, todo proceso se resuelve en el nivel de las teorías básicas de la física (p. 343). Desde una perspectiva más amplia de los tipos de causalidad, si bien aún se mantiene la influencia causal de los niveles de organización inferiores sobre los superiores, en un sentido especial estos últimos también poseen poder causal sobre los primeros. Así, respecto de un organismo cualquiera, dada la continuidad material entre la física y la biología, aquel estará necesariamente determinado por las leyes que rigen los fenómenos del nivel inmediato inferior. Sin embargo, en tanto que el todo pertenece al orden de lo biológico, se rige, como tal, de acuerdo con patrones que pueden entenderse en términos de la causalidad formal y funcional (por ejemplo, mantener la estabilidad del sistema, en este caso, mantener vivo al organismo, determina las funciones específicas de sus componentes), patrones que enmarcan la acción de las propiedades del nivel organizativo inferior.

Estas consideraciones sobre los tipos de causalidad ayudan a mantener la inteligibilidad del poder causal de las propiedades supervenientes frente a las objeciones reduccionistas. No obstante, El-Hani y Pereira van más lejos en su argumentación, y sostienen que las propiedades y principios que rigen las relaciones de los niveles organizativos inferiores solo resultan explicativas respecto de las entidades que conforman los niveles superiores si es que se las lee en virtud de estas últimas. En el caso de la biología, los autores llaman a esta noción 'significatividad biológica' (p. 344), y es su principal argumento contra el reduccionismo. El argumento se funda en una idea que, a pesar de que pareciera ser de sentido común, el reduccionismo olvida fácilmente. La pertinencia, en el área de, por ejemplo, la biología, del recurso a propiedades de sistemas más básicos solamente se puede determinar a la luz de las propiedades de biológicas mismas. Resulta absolutamente intuitiva, como señalan El-Hani y Pereira, la diferencia entre «una masa aleatoria de moléculas y un organismo» (p. 344). Si se va a recurrir a la reducción de algún fenómeno biológico a uno molecular, la reducción solo tiene sentido si se le asigna un puesto específico dentro del ordenamiento general de las teorías biológicas.

Los mismos autores son conscientes de que su argumento final se inscribe dentro de una postura metafilosófica particular (p. 343). La posibilidad o imposibilidad de la reducción no sería, como admiten junto con Davidson y Rorty, una disputa ontológica, sino, más bien, sobre el estatus de posibles descripciones de la realidad (pp. 343-345 y nn. 35, 36, 40 y 41).

Según Peter Menzies (2010), la mayor parte de las opiniones tradicionales en torno del estatus epistemológico de la psicología intencional se pueden dividir en dos grandes grupos. Por una parte está la tradición que el autor denomina *naturalismo científico*, que postula que la explicación del comportamiento de los agentes depende de esquemas nomológico-deductivos. De esta manera, en tanto intérpretes aplicamos, normalmente de manera no explícita, leyes que gobiernan los nexos entre los estados mentales de los agentes y las acciones que observamos. Dichos nexos son, por lo general, pensados como nexos causales, tal y como ocurre en las ciencias de la naturaleza. De hecho, desde esta perspectiva suele asumirse que los nexos causales son parte del ámbito de lo físico, por lo cual se concluye que las relaciones que explican las leyes de la psicología intencional también son relaciones causales físicas.

De otra parte, se encuentra la tradición que el autor denomina *racionalismo kantiano*. Aquí se propone que los principios de la psicología intencional cumplen un rol normativo, ya no descriptivo como en el caso anterior. Es decir, dichos principios explican el comportamiento de los agentes según lo que solemos esperar a partir de nuestra concepción general de la racionalidad. Los autores que pertenecen a este grupo tienden a distinguir entre *racionalizar* una acción (hacerla inteligible en tanto acto de un agente) y *causarla*. En su concepción, los contenidos intencionales atribuidos en las explicaciones psicológicas de la acción solo hacen lo primero, racionalizan, mas no causan, pues, al igual que quienes pertenecen al otro gran grupo, consideran que las causas siempre pertenecen al dominio de lo meramente físico.

Como señala Menzies, uno de los principales argumentos desarrollados por filósofos desde mediados del siglo XX con el fin de separar de una vez por todas la explicación causal de la psicológica (y que, por ende, sirve para justificar la oposición entre naturalistas científicos y racionalistas kantianos), es el llamado «argumento de la conexión lógica» (2010, pp. 145-149). La estrategia consiste en resaltar dos rasgos interrelacionados de las explicaciones de la psicología intencional. Las generalizaciones mediante las cuales explicamos la conducta de un ente racional no parecen ser empíricas, sino únicamente conceptuales. Pensamos los actos de alguien de acuerdo con lo que idealmente concebimos como lo más racional y le atribuimos los estados mentales correspondientes al agente de las acciones. Sin embargo, por más que el agente efectivamente haya realizado los juicios que le atribuimos de acuerdo con nuestros estándares de racionalidad, aquel podría actuar de otro modo. Los casos de debilidad de la voluntad en que los agentes no parecen decidir racionalmente son abundantes entre nosotros y ello parece mostrar que las generalizaciones de la psicología intencional no son verdades empíricas. El otro aspecto de la explicación intencional que incentiva el escepticismo con relación al poder causal de las razones queda evidenciado en las expresiones que utilizamos normalmente en semejantes contextos. Usamos, por ejemplo,

palabras como el auxiliar verbal «debería» cuando pensamos la opción según la cual el agente cumpliría mejor con los estándares de racionalidad. Como dijimos arriba, aquí se hace manifiesto el carácter normativo de la explicación intencional.

Aparte de los dos grupos tradicionalmente mayoritarios en cuanto a la discusión sobre la naturaleza de la explicación intencional, existe una tercera vía, cuyo principal representante es Donald Davidson. Menzies denomina a esta opción «compatibilismo», y es la que se propone defender, aunque con ciertas diferencias respecto de la versión davidsoniana, a la que denomina «compatibilismo de dos niveles» (pp. 145-146 y 149-153). En dos artículos clásicos —«Actions, Reasons and Causes» (2001a) y «Mental Events» (2001b)—, Davidson intenta conciliar lo que el argumento de la conexión lógica pretende demostrar que es irreconciliable: primero, que las razones atribuidas a un agente con el fin de racionalizar sus acciones son, al mismo tiempo, las causas de sus acciones; luego, que este poder causal de las razones es compatible con una concepción nomológica de la causalidad (como la que cree el propio Davidson que pertenece a las ciencias naturales) y con la irreductibilidad de la explicación psicológica a la meramente física.

Resulta esencial para la argumentación de Davidson su concepción de la causalidad como una relación extensional entre eventos. Así, un mismo nexo causal entre dos sucesos puede ser descrito en términos mentales (empleando, por ejemplo, los llamados verbos mentales, como *creer*, *pensar*, *desear*, etcétera) y satisfacer los principios de la explicación psicológico-intencional, y, además, ser descrito en términos físicos y satisfacer las explicaciones causales científicas construidas sobre la base de leyes. Pero, como también hacen muchos otros intérpretes de la propuesta davidsoniana, Menzies considera que esta salida trae consigo un costo elevado para la psicología. El problema radica en que Davidson acepta como una de sus premisas iniciales que todo nexo causal debe instanciar, al menos en una de sus descripciones posibles, leyes que lo gobiernen. Esto parece obligar a la psicología intencional a que recurra a dos tipos de explicaciones distintas: las racionalizaciones, formuladas en términos, como vimos, propiamente psicológicos, y las explicaciones causales, que únicamente cumplen la exigencia nomológica en términos físicos.

Menzies argumenta, contra lo expuesto arriba sobre Davidson, que en la realidad la psicología únicamente requiere de un mismo nivel de explicación para cubrir tanto la función de racionalizar como el esclarecimiento del nexo causal. Tanto quienes aceptan el argumento de la conexión lógica como el propio Davidson habrían cometido el mismo error, o sea, asumir que las teorías científicas contienen explicaciones causales en virtud de explicaciones nomológico-deductivas. En lugar de ello, pueden pensarse, por ejemplo, las leyes estrictas de la física, no como representaciones de regularidades empíricas, sino como los elementos que estructuran modelos ideales

que cumplen objetivos explicativos específicos dentro de las ciencias (Menziez, 2010, pp. 154-159). En otras palabras, el poder explicativo de las teorías científicas no descansaría en si las leyes que las componen son verdaderas o falsas respecto del mundo. Su poder explicativo depende de cuán eficientemente se pueden emplear como modelos que instancian sistemas reales (2010, p. 156).

Las teorías de la psicología intencional no se distinguirían de lo que hemos dicho sobre las condiciones de la capacidad explicativa de las teorías científicas en general. Por ende, semejantes teorías son, también, modelos. En este caso, se trata de modelos contruidos sobre la base de los principios que creemos componen la racionalidad de agentes ideales. El resultado de la aplicación de estos modelos es siempre contingente. Esto último resulta importante si recordamos que, de acuerdo con Menziez, la diferencia principal entre las explicaciones psicológicas y aquellas de las otras ciencias radica en que las generalizaciones de la psicología no parecen ser tan estrictas como lo requiere la concepción nomológico-deductiva de la causalidad. De ahí que se requieran dos niveles de explicación, uno que racionalice y otro que apele a leyes estrictas que sí satisfagan lo requerido para el esclarecimiento de los nexos causales. No solo esto debería poder ser superado por la idea de la explicación científica como la aplicación de modelos ideales, también el segundo punto del argumento de la conexión lógica debería superarse. Dependiendo de para qué se le requiera, un mismo modelo psicológico podría servirnos en la explicación causal, cumpliendo un papel descriptivo, y en consideraciones normativas sobre la conducta. La diferencia en la aplicación del modelo en uno y otro caso depende de su semejanza con el sistema real. Si el modelo difiere del sistema real en la acción resultante, diremos, normativamente, que el agente *debió* actuar de otra manera.

1.5.3. Evolución de la agencia

Tal vez sea posible trazar el camino por donde pueda empezar a explicarse el paso evolutivo de la mera detección y respuesta frente a estímulos ambientales, a la agencia y la preferencia, es decir, a representar y decidir. Kim Sterelny (1999) propone una hipótesis que intenta cumplir con ello. En primer lugar, debemos considerar que existen buenas razones para cuestionar la imagen clásica de la inteligencia. Se trata de la idea según la cual el comportamiento inteligente es producto de la capacidad del agente para representarse el mundo de la manera más completa posible. El problema radica en que, dada la complejidad de cualquier entorno, resulta difícil imaginar que una mente sea capaz de mantener una representación adecuada del mismo y, a la vez, seleccione sus factores más relevantes para coordinar el comportamiento (1999, pp. 203-204). Por ello, hay quienes se inclinan en favor de pensar la inteligencia como una facultad situada, o sea, como un sistema que surge de la interacción entre

módulos especializados y ciertos elementos del entorno. Naturalmente, se trata de una visión emergentista de la inteligencia. Esta última, en tanto todo compuesto por niveles inferiores de organización de los módulos específicos, poseería propiedades únicas. No tendría caso, entonces, tratar de comprender la inteligencia así descrita analizando por separado las funciones de sus componentes (p. 205).

Sterelny sostiene que la necesidad de determinar los datos del entorno importantes para la supervivencia del organismo, son una forma de presión evolutiva que desemboca en la selección, tanto de tipos de comportamiento que se explican únicamente como producto de una relación entre ciertos dispositivos de detección y el entorno, como de la representación. Muchas veces ocurre que los organismos deben hacerle frente a lo que el autor llama «polución epistémica» (p. 208), esto es, a la emisión de datos equívocos que exigen estrategias de desambiguación. La capacidad representacional probaría ser especialmente útil para ello. Por ejemplo, medios hostiles, como aquellos que cuentan con depredadores altamente especializados, favorecerían la selección de la representación.

La representación no basta para explicar la agencia. Para elegir, también se necesita la capacidad de desear. La complejidad del entorno puede seleccionar la representación. Pero, cuando un agente elige, lo hace encarando la complejidad de sus posibles respuestas frente a los estímulos del entorno. Otra característica fundamental de los seres que poseen agencia es que estos son capaces de representar sus propios estados internos, que actúan como motivaciones para las posibles acciones. A todo ello debe agregársele que los seres en cuestión deben ser capaces de representarse el entorno como debería ser, no tan solo como es (p. 211).

Para explicar el paso de la mera detección de los estados internos de los organismos a su representación, Sterelny sugiere que consideremos el caso de una rata que, frente a la deshidratación, comienza a beber agua (pp. 211-212). La rehidratación de las células demora entre 10 y 15 minutos, pero el animal dejará de beber antes de ese tiempo. Aquí el estímulo que pone en marcha la acción no puede ser el mismo que la detiene. Resulta factible, entonces, que en un sistema semejante, la representación del estado interno ayude a coordinar la acción al permitir el rastreo de dos tipos de estímulo diferentes. No obstante, un ejemplo como el de la rata es insuficiente, pues ambos estímulos podrían ser internos. El autor piensa que el caso paradigmático de situación en la que se presenta un estímulo externo para cesar la acción o controlarla es el enfrentamiento con depredadores.

Sterelny también sostiene que puede haber otro tipo de factores que complejicen los sistemas de control del comportamiento. A veces el comportamiento mismo se complejiza, como en el caso de los braquiadores pesados (los grandes simios) que vimos en secciones anteriores respecto de la hipótesis de Byrne (1997) sobre la inteligencia

técnica en primates. Inclusive, una posible acción podría interferir con el desarrollo de otra, y el animal debería poder controlar su propio desempeño mediante estrategias en las cuales las acciones se ordenen jerárquicamente. El propio Byrne se refería a esto como una cierta capacidad de planificación y, por ello, no resulta extraño que Sterelny encuentre atractiva la hipótesis de aquel.

Aparentemente, entonces, frente a opciones cada vez más complejas de comportamiento, «el control se vuelve menos discreto, menos modular» (Sterelny, 1999, p. 214). Otra conclusión importante que se extrae de los argumentos de Sterelny es que la preferencia depende de la representación, pero no se puede descartar la posibilidad de que exista representación sin preferencia, pues su tesis sobre el desarrollo de la segunda toma como un hecho inicial la primera. Sin embargo, el autor no profundiza mucho más en la relación. De la misma manera, con lo dicho hasta aquí solo nos alcanza para afirmar que, aparentemente, la complejidad de los estímulos de ciertos entornos podría seleccionar la capacidad representacional y, a su vez, en algunos casos ello conduciría al desarrollo de la agencia. Pero el vínculo entre ambas capacidades cognitivas aún no ha sido esclarecido exhaustivamente.

1.6. Síntesis

Como se vio al principio de este capítulo, Darwin no solo estableció las bases del paradigma que hasta hoy rige la biología evolutiva, sino que, más específicamente, ya en su trabajo se encuentran las primeras pistas para el estudio filogenético de la cognición social humana. Siguiendo esta misma línea de investigación evolutiva, desde las últimas décadas del siglo pasado, los avances en neurociencias y en primatología les han brindado sustento empírico a las teorías de la psicología y la filosofía acerca de los orígenes de nuestras facultades cognitivas; trabajos como los de Panksepp (2007), Premack y Woodruff (1978) y Premack (1988) son claras muestras de ello.

La llamada «hipótesis de la inteligencia maquiavélica» es la teoría más aceptada sobre las presiones selectivas que han determinado la evolución de la cognición social. Sin embargo, a pesar de que, como se explica mediante dicha teoría, la competencia entre individuos ha sido crucial en la selección de muchas de nuestras facultades cognitivas más distintivas, la tesis requiere de algunos complementos. De esta manera, y continuando con una tradición iniciada por Darwin, autores como Dunbar (2000) destacan el rol que ciertos mecanismos de cooperación han jugado en la evolución de nuestra capacidad para atribuirles varios niveles de intencionalidad a otros miembros de nuestra especie (por ello, actualmente se prefiere sustituir el término *inteligencia maquiavélica* por el más general de *inteligencia social*). Asimismo, existen teorías complementarias que toman como punto de partida algún mecanismo de interacción social; este es el caso de Byrne (1997), cuya tesis revisamos en este mismo capítulo.

Finalmente, examinamos cómo hoy en día algunos problemas clásicos de la filosofía y la psicología son abordados desde las teorías filogenéticas de la ciencia cognitiva. Un ejemplo de ello es el debate actual respecto de nuestro acceso consciente a las mentes de los demás y a nuestras propias mentes. Para unos, como Goldman (2006), el conocimiento que alcanzamos de la mente de los demás depende de nuestra capacidad para simular en nosotros mismos los posibles estados mentales ajenos. Para otros, como Carruthers (2009), el camino es el inverso: el conocimiento que tenemos de nuestros propios estados mentales es un producto colateral de nuestra actividad como lectores de otras mentes. En buena cuenta, el debate se libra en el terreno filogenético. Otro ejemplo expuesto arriba es el viejo problema de la aparente incompatibilidad entre la imagen determinista de la naturaleza, ampliamente extendida en el pensamiento científico, y la idea de sentido común de la eficacia causal de la mente (es decir, la idea de agencia). Si una teoría naturalista de la evolución de la agencia, como la propuesta por Sterelny (1999), es exitosa, quienes aquí vimos que abogan por la irreductibilidad de conceptos psicológicos de sentido común (El-Hani & Pereira, 1999; Menzies, 2010) verían sus posiciones reforzadas.

2. EL DESARROLLO DE LA COGNICIÓN SOCIAL

Hasta aquí hemos presentado la cognición social desde el punto de vista de la evolución, es decir, en términos filogenéticos. A continuación, intentaremos relacionar esta información con la que se tiene a nivel del desarrollo del individuo, es decir, desde una perspectiva ontogenética. Un puente interesante lo proporciona la psicología evolucionista del desarrollo, que trataremos a continuación. Luego detallaremos algunos hitos en el desarrollo esenciales para la cognición social y finalizaremos profundizando en la teoría de la mentalización, que forma parte de aquellas propuestas que enfatizan que es a partir de las relaciones con otras personas que el individuo desarrolla las habilidades necesarias para la cognición social.

2.1. El desarrollo desde el punto de vista evolucionista

La psicología evolucionista del desarrollo (PED) es un campo que se interesa por el desarrollo en la infancia, niñez y adolescencia desde una postura evolucionista. Asimismo, analiza las conductas parentales que evolucionaron para brindar el cuidado requerido durante dichas etapas del desarrollo. Busca describir y comprender aquellas conductas y rasgos psicológicos del cuidado parental que son adaptaciones diseñadas para incrementar la supervivencia en la adultez. Ello implica «[...] el estudio de los mecanismos genéticos y ambientales que subyacen al desarrollo universal de las competencias cognitivas y sociales, y la epigénesis evolucionada

(interacciones genes-ambiente) que adaptan esas competencias a las condiciones locales» (Hernández & Bjorklund, 2008, p. 182). Es decir, la PED se pregunta cómo los mecanismos heredados se desarrollan y se expresan en la adultez a través de la interacción de los individuos con su ambiente. Es necesario realizar un análisis evolucionista y epigenético para entender el desarrollo humano en nuestro mundo actual. Con esta finalidad, la psicología evolucionista del desarrollo ha recogido los aportes tanto de la psicología evolucionista como de la psicología del desarrollo.

La psicología evolucionista (PE) es un modelo que intenta unificar e integrar distintas áreas y corrientes de la psicología tradicional (Dunbar & Barrett, 2007). Es una psicología que recoge información de la biología evolucionista y postula que la arquitectura de nuestra mente y, por tanto, la conducta humana, son el resultado del proceso de evolución (Barkow, Cosmides & Tooby, 1992). Ahora bien, el diseño de nuestra mente, producto del proceso de selección, no está orientado a solucionar los problemas del humano actual, está dirigido a resolver los problemas adaptativos que los individuos tuvieron que afrontar como cazadores-recolectores. Cabe señalar que un problema adaptativo es aquel cuya solución puede afectar la reproducción, aunque sea de manera lejana o indirecta. Algunos problemas adaptativos son evitar el incesto, identificar plantas y animales, cazar animales, adquirir la gramática, responder a llamadas de alarma, detectar cuando los niños requieren ayuda, querer atender las demandas de los niños, seleccionar parejas, ser atractivos para las parejas, interpretar situaciones sociales, ayudar a familiares, mantener relaciones de amistad, reconocer emociones, cooperar, entre otros (Buller, 2005). La tarea de los psicólogos evolucionistas está orientada a encontrar las conexiones necesarias entre los problemas adaptativos y la estructura de los mecanismos que evolucionaron para resolverlos (Barkow, Cosmides & Tooby, 1992). Su finalidad última es descubrir, describir y explicar la naturaleza de dichos mecanismos puesto que para ellos la evolución no selecciona conductas sino los mecanismos que generan la conducta.

Por su parte, la psicología del desarrollo o psicología evolutiva se ocupa en describir y explicar los cambios en la estructura física, el comportamiento y el funcionamiento mental que surgen a lo largo de todo el ciclo vital, es decir, desde la vida intrauterina hasta la muerte (Stassen, 2008). Compartimos la propuesta de que los seres humanos se desarrollan de una manera dinámica, continua y en constante interacción con su ambiente, lo cual incluye las relaciones interpersonales e intergrupales (Cairns, Elder & Costello, 1996).

A partir de la integración de estas dos fuentes de información, la PED ha elaborado una serie de planteamientos básicos: todos los mecanismos evolucionados se desarrollan a partir de continuas interacciones entre los genes y el ambiente, en una influencia mutua; la selección natural está presente en todos los periodos del

desarrollo, no únicamente durante la adultez; ciertas características de los infantes y los niños fueron seleccionadas para ejercer funciones adaptativas en momentos específicos del desarrollo y no como preparaciones para la adultez; las adaptaciones de los niños presentan cierto grado de plasticidad o flexibilidad, lo que implica capacidad para cambiar en respuesta a distintas condiciones ecológicas y ambientales; los individuos requieren de una niñez extendida para comprender las complejidades de las comunidades sociales humanas; y, finalmente, la PED señala que tanto los mecanismos cognitivos de dominio específico como los de dominio general fueron modelados por la selección natural.

Del mismo modo, la PED plantea la existencia de tres tipos de adaptaciones. Las adaptaciones ontogenéticas son aquellas para periodos particulares del desarrollo que desaparecen cuando ya no son necesarias. Por ejemplo, el cordón umbilical y la imitación de gestos faciales en niños menores de dos años. Las adaptaciones diferidas (*deferred*) funcionan para preparar a los niños para temas específicos de la adultez. Por ejemplo, la diferencia en los estilos de juego entre niños y niñas. Por último, las adaptaciones condicionales, que aparecen en los momentos en que un individuo responde a las condiciones de su entorno regulando su conducta de manera anticipada a las condiciones de la vida futura. Por ejemplo, el modo de crianza en la infancia predice el *timing* de la pubertad y la estrategia de elección de pareja en la adultez.

Dentro de este modelo, los autores Flinn y Ward (2005) se interesaron por encontrar un vínculo entre la niñez y la mente social. Ellos consideran que la competitividad social ha sido la presión de selección principal que ha dado forma a la combinación humana de ser físicamente altricial, pero mental y lingüísticamente precocial durante la infancia, la niñez y la adolescencia. Esta combinación es posible gracias al cuidado parental y familiar extendido.

Debido a que los humanos se volvieron tan ecológicamente dominantes, ellos mismos se tornaron en su principal fuerza hostil de la naturaleza. Así, aquellos individuos más diferenciados cognitivamente, social y conductualmente tendrían mayor oportunidad para ejercer control sobre otros individuos y tener mayor acceso a los recursos de su medio, todo lo cual estaría al servicio de la supervivencia y la reproducción. Como consecuencia, para los autores se trata de una presión selectiva esencial que moldeó los principales procesos mentales de los humanos como, por ejemplo, la teoría de la mente y el lenguaje.

Ahora bien, ¿por qué ello daría lugar a una niñez prolongada? La respuesta está en que las relaciones sociales son inmensamente complejas y cambiantes, por lo cual se requiere del desarrollo de herramientas con las mismas características; ello toma tiempo. El aprendizaje, la práctica y la experiencia son indispensables para desarrollar las herramientas sociales necesarias y así obtener éxito.

De otro lado, Flinn y Ward (2005) señalan que la variedad en las relaciones interpersonales favorece la evolución de sistemas cerebrales y cognitivos más generales, capaces de integrar el procesamiento realizado por mecanismos de dominio específico. Además, hacen hincapié en la plasticidad de los sistemas neuronales, lo cual permite que el niño se adapte a su comunidad local como, por ejemplo, a su lenguaje. Cabe mencionar que el lenguaje permite un mecanismo nuevo para el aprendizaje social y la cultura; el patrón recursivo y la representación simbólica facilitan un procesamiento de información abierto, creativo y flexible.

Finalmente, como se mencionó al inicio, los padres y otros familiares juegan un rol importantísimo en el desarrollo mental del niño pues son ellos quienes, por un lado, brindan una garantía de supervivencia mientras el niño va adquiriendo las habilidades sociales necesarias y, por otro lado, constituyen puntos de referencia confiables que influyen de manera directa en el aprendizaje social del niño.

2.2. Hitos del desarrollo de la cognición social

En esta sección, señalaremos algunos logros del desarrollo que se relacionan directamente con la cognición social (Thompson, 2006). De manera general, a lo largo del primer año de vida, los infantes aprenden a distinguir el mundo social del mundo inanimado gracias a las preferencias perceptuales que los orientan hacia eventos sociales. Asimismo, poco tiempo después del nacimiento, empiezan a diferenciar entre lo interno y lo externo. Según la teoría perceptual contemporánea, ello sucede mediante la percepción, ya que esta permite distinguir entre la acción iniciada por la propia persona y aquella generada en el entorno. A partir de este punto, poco a poco, el individuo va adquiriendo conciencia de sí mismo.

Entre los dos y tres meses, ya son capaces de distinguir características faciales más detalladas. Además, comprenden que sus acciones generan una reacción en las demás personas y que esa reacción, a su vez, puede producir excitación y ánimo positivo en ellos mismos. No solo eso, sino que esperan que las personas —familiares y desconocidas— interactúen con ellos. Es importante recalcar que las diversas respuestas de los adultos tienen efectos en las reacciones emocionales y sociales de los infantes y que ello se generaliza hacia otras personas. Es decir, se van generando pautas o modelos de relaciones que marcarán una tendencia en el infante al encontrarse con otros sujetos. Hacia los tres meses, aparecen los juegos cara a cara con cuidadores conocidos, una de las formas más tempranas de interacción social focalizada y de aprendizaje social. En este intercambio socioemocional, el bebé descubre que los encuentros sociales son dinámicos y cambiantes, no siempre claros y previsibles. Por otro lado, en esta etapa, desarrollan la habilidad para representar los actos de los demás como si fueran ellos los que también podrían realizarlos. Por último, otra capacidad que surge

en este momento es aquella que permite discriminar entre las distintas expresiones emocionales de la madre, lo que hace que expresen desconcierto cuando no hay congruencia entre las expresiones faciales y vocales de la madre.

Hacia los seis meses, los bebés ya saben perfectamente que la conducta de los cuidadores para calmarlos les genera una sensación de alivio. Con ese conocimiento, se producen expectativas sobre aquellas personas, tanto así que muchos infantes se empiezan a calmar con solo escuchar los pasos de un adulto acercándose.

Alrededor de los siete meses, gracias al desarrollo motor, los bebés adquieren una mayor independencia; son capaces de acercarse —o desplazarse si gatean— y explorar el entorno sin ayuda de otros. Se producen grandes avances en el sistema perceptual y, también, en el intercambio social y con el mundo. Los bebés progresan en la comunicación referencial, la comprensión de la relación entre los medios y el fin, y la interacción social a distancia. La mayor movilidad e independencia del infante permite la expresión de sus deseos y facilita una mayor conciencia de las intenciones de las demás personas. Al mismo tiempo, los cuidadores empiezan a poner límites y prohibiciones más claros para proteger al bebé.

Entre los nueve meses y el año de vida, los infantes empiezan a entender las conductas de los demás como intencionales y dirigidas hacia una meta, ya que son interpretadas en asociación a los objetos hacia los cuales están dirigidas (Thompson, 2006). Cerca al año, el infante adquiere la capacidad para la atención conjunta, la cual consiste en una coordinación de la atención entre dos individuos y un objeto o evento, es decir, una interacción triádica. Ahora bien, no se trata solo de dos personas dirigiendo su atención hacia un mismo punto, sino que ambos han de ser conscientes de que están atendiendo a algo en común; es decir, se requiere de la conciencia de experiencia compartida. Las conductas más representativas de la atención conjunta son los gestos para señalar y las respuestas frente a los gestos para señalar de las otras personas. Así, se observan gestos protodeclarativos para dirigir la atención de otra persona —como cuando el infante alterna la mirada entre la cara del adulto y un objeto mientras lo señala— y gestos protoimperativos que transmiten un pedido —por ejemplo, cuando el infante intenta agarrar algo alternando la mirada entre el adulto y el objeto que intenta agarrar— (Carpendale & Lewis, 2010).

La habilidad para la atención conjunta es considerada fundamental por muchos pues constituye la base para una mayor interacción comunicativa, para la adquisición del lenguaje y para la participación en procesos culturales. En este sentido, la atención conjunta sería un aspecto esencial para el desarrollo de la cognición social. De hecho, en algunos estudios longitudinales, se ha encontrado que la atención conjunta correlaciona con la posterior competencia social, y el desarrollo cognitivo y del lenguaje (2010).

Para algunos autores como Trevarthen (Beebe, Sorter, Rustin & Knoblauch, 2003) las personas venimos al mundo con una intersubjetividad innata que nos permite captar intuitivamente —sin requerir del lenguaje— las motivaciones y sentimientos del otro; venimos con un potencial para comunicarnos. Trevarthen propone una secuencia para el desarrollo de este potencial. En primer lugar, nacemos con una *intersubjetividad primaria* que nos permite una comunicación diádica basada en la forma, la sincronía y la intensidad. Aquí se ubica la protoconversación. Luego, en una especie de paso intermedio, se ubican los juegos que, paulatinamente, irán dando paso a la *intersubjetividad secundaria*. Aquí se ubica la atención conjunta descrita más arriba y se empieza a formar una teoría de la mente en tanto empieza a surgir un pensamiento simbólico.

Al año de edad, ya se aprecia con bastante claridad el uso de la referencia social, que permite comprender aquellas situaciones que resultan ambiguas a partir de una lectura de las señales emocionales de otros. Esta información captada influye directamente en las propias respuestas del infante, tanto emocionales como conductuales (Thompson, 2006).

A los dieciocho meses, las inferencias que los infantes hacen sobre las intenciones de los adultos conforman la base para el aprendizaje inicial de palabras. Asimismo, en el desarrollo del lenguaje, se observa un esfuerzo por comprender las normas del entorno social. Adicionalmente adquieren la conciencia de que los deseos de las demás personas pueden ser distintos de los propios y comprenden bastante más la relación entre el deseo y la emoción: cuando los deseos se cumplen uno está contento. Alrededor de esta edad, los infantes son capaces de reconocerse a sí mismos en el espejo y se avergüenzan si su apariencia física no se adecua a lo que se espera (por ejemplo, si tienen pintada la frente).

A los dos años, los niños son capaces de representarse eventos, como las rutinas cotidianas, y de ese modo tener expectativas sobre el futuro. La organización, la estructura y el contenido de los eventos no solo se elabora a partir de lo prelingüístico sino también por la estructura verbal que se le da en el intercambio padre-hijo. La representación de eventos en general es la base para la comprensión de eventos sociales. Por otro lado, los niños a esta edad son capaces de hacer referencias verbales sobre las emociones, sus causas e, incluso, los esfuerzos para su regulación.

A los tres años, los niños saben que las emociones se vinculan a creencias y expectativas, y son conscientes de que las creencias pueden no ser consistentes con la realidad. Es decir, comprenden que, frente a una misma experiencia, las personas pueden generar diversas interpretaciones.

A los cinco años de edad, los niños son capaces de entender que los sentimientos pueden ser reactivados a partir de la conexión entre un evento actual y recuerdos

del pasado. Por ejemplo, si vemos a alguien que se parece a un familiar fallecido podría producirnos pena. Asimismo, comienzan a percibir a los otros en términos de rasgos psicológicos y pueden predecir el comportamiento futuro a partir de las características que han inferido de una persona (2006).

Hasta aquí, hemos presentado algunos aspectos importantes de la primera infancia que permiten apreciar el modo en que las distintas habilidades sociales van surgiendo y que permitirán —si los logros son óptimos— una comprensión y manejo adecuados del mundo social.

2.3. La mentalización. Una perspectiva relacional

En la actualidad, no hay duda de que la cognición social es el resultado de una combinación entre lo biológico y lo ambiental. Las diferencias entre teorías radican, más bien, en el peso que se le otorga a cada aspecto. Algunas teorías, aunque no niegan la influencia del entorno social, se concentran fundamentalmente en los aspectos biológicos; estudian la cognición social desde las capacidades mentales de los individuos. La teoría de la mente se ubicaría aquí, por ejemplo. Otras, como la que presentaremos, adoptan un enfoque relacional para dar cuenta de la cognición social. Desde esta aproximación, se asume que es a partir de los procesos sociales que se desarrolla la mente.

La perspectiva relacional parte de la premisa de que el bebé, al nacer, no se diferencia de su cuidador. Esto es, no puede delimitar con claridad que él es uno y el cuidador otro. A partir de las acciones en el mundo de ambos, este sistema unificado irá perdiendo fuerza para dar paso a la diferenciación. Recién en este punto el bebé podrá darse cuenta de que el cuidador y él son dos, y no uno. Así, el desarrollo de la comprensión social y de uno mismo parte de la relación entre dos individuos. Por tanto, para entender cómo surge la mente y, por ende, la cognición social, es fundamental atender el aspecto relacional (Carpendale & Lewis, 2010).

La teoría del apego se ubica dentro de esta propuesta y considera que la diferenciación se logra a partir de ciertas relaciones de apego fundamentales. Esta teoría plantea que existe un sistema cerebral que orienta al bebé a buscar cercanía y comunicación con sus cuidadores. En principio, ello cumple un rol importante para la supervivencia, pero este sistema es más complejo aún. A nivel mental, el cerebro poco desarrollado del bebé, a partir del encuentro interpersonal, utiliza las funciones ya maduras del cerebro de sus cuidadores para organizar sus propios procesos mentales. A partir de estas relaciones emocionales se desarrolla una serie de funciones mentales importantísimas como la memoria, la narración, las emociones, las representaciones y los estados mentales (Siegel, 1999). Así, el apego podría funcionar como una base sobre la cual se desarrollaría la mente.

Para que esto se logre de manera óptima, aquello que se denomina *apego seguro*, es esencial que los cuidadores sean sensibles a las señales del bebé, pues de ese modo podrán amplificar las respuestas emocionales positivas del bebé y modular las negativas. El cuidador ayuda al bebé a disminuir la intensidad de emociones desagradables como el miedo, la tristeza o la ansiedad, con lo cual se tranquiliza, y, a la vez, le ofrece consuelo frente a ese malestar, es decir, seguridad. A partir de las repeticiones de este patrón de intercambio se van a ir generando modelos que el bebé luego interiorizará y hará propios. Cuando esto suceda, será capaz de proporcionarse la tranquilidad y seguridad por sí mismo, sin necesidad del cuidador. Varios estudios sobre este tema han encontrado que los modelos generados a partir del sistema de apego se asocian con los procesos particulares de la regulación emocional, las relaciones sociales, la disponibilidad de la memoria autobiográfica y el desarrollo de la autorreflexión y la narración (1999). Por otro lado, también se ha encontrado en investigaciones una asociación entre el apego seguro y el desarrollo social de los niños, la cual incluye el hallazgo de que los niños con apego seguro tienen más probabilidades de pasar la prueba de la falsa creencia a alrededor de los cuatro años que los niños con apego inseguro (Carpendale & Lewis, 2010).

La mentalización se encuentra muy próxima a la teoría de apego, como se verá a continuación. La mentalización o función reflexiva del *self* es la habilidad que permite a las personas percibir e interpretar su propia conducta y la de los otros en términos de estados mentales intencionales, e imaginar lo que los demás sienten y piensan (Busch, 2008). Dentro de los estados mentales se encuentran los deseos, los sentimientos, las creencias, las fantasías, las expectativas, entre otros. Esta habilidad incluye, en realidad, una serie de capacidades tanto explícitas como implícitas, es decir, que se ubican en niveles conscientes y no conscientes o automáticos. Dentro del primer grupo se ubican, por ejemplo, la interpretación verbal de un evento o la elaboración de una explicación, mientras que en el segundo se encuentran la intuición o la sintonía afectiva entre madre e hijo (Fonagy, 2008).

Hace poco, influidos por la psicología evolucionista, Fonagy y Target (1997) han redefinido la función reflexiva del *self* por la función interpretativa interpersonal del *self* (IIF, por sus siglas en inglés), en un intento por resaltar el componente intersubjetivo de esta capacidad y su impacto tanto en la evolución como en el desarrollo de la vida mental y social de los individuos (Bouchard & Lecours, 2008).

La IIF permite procesar las experiencias nuevas e interpretarlas en términos psicológicos a partir de la atribución de intencionalidad. Esta función presenta cuatro elementos que se desarrollan de manera sucesiva en el contexto del vínculo de apego: (a) el reconocimiento y regulación de afectos, (b) la regulación de los estados de activación internos, (c) la regulación cognitiva —que implica el control de la atención

y la conceptualización de los estados mentales— y (d) las capacidades específicas de la mentalización. Cabe resaltar que la habilidad para mentalizar se concretiza entre los tres y los seis años.

Fonagy postula, al igual que el biólogo evolucionista Alexander (1989, citado por Fonagy, 2008), que las habilidades mentales más sofisticadas, como la inteligencia social, fueron impulsadas por la competencia entre los seres humanos. Luego de alcanzar cierto grado de dominio sobre el ambiente natural, las personas nos convertimos en nuestra principal amenaza y, debido a ello, tuvimos que aprender a cooperar entre nosotros. Competir con otros exige entenderlos y predecir sus intenciones. De este modo, las capacidades para construir y manipular situaciones en la mente sobre los pensamientos y sentimientos se tornaron en una ventaja reproductiva.

Para Fonagy (2008), los vínculos de apego en los seres humanos generan un ambiente seguro para el adecuado desarrollo del cerebro social. Las capacidades de mentalización, al igual que las otras habilidades cognitivas necesarias para su crecimiento, evolucionan a partir de los intercambios sociales con los cuidadores principales. Como mencionamos anteriormente, el crecimiento y la complejización de las habilidades sociales evolucionaron junto con otros factores del desarrollo, como el aumento de la indefensión en los infantes y la prolongación de la infancia, y la aparición de formas de paternaje más sofisticadas.

A continuación, se presentarán los hitos ontogenéticos, agrupados en cuatro etapas secuenciales que delinearán el desarrollo de la capacidad para la mentalización (2008).

La primera etapa, en la cual se aprecia una *agencia teleológica*, está conformada por la segunda mitad del año de vida. Durante este periodo el infante logra captar e identificar la relación causa-efecto. Espera desenlaces determinados para acciones conocidas y reconoce que estas son medios para lograr fines específicos. La agencia es entendida en términos de la elección de los medios más eficaces para lograr el objetivo esperado. Debido a ello, el infante tiene la expectativa de que los otros agentes llevarán a cabo las acciones que sean las más eficaces para alcanzar una meta particular. Todavía no es capaz de atribuir estados mentales como motivadores de la conducta ni de considerar otras perspectivas distintas de la suya. Al acercarse al primer año de vida, las habilidades comunicativas del infante se tornan más elaboradas; empieza a hacer uso de la referencia social, es decir, busca el rostro del adulto para obtener información sobre si debe acercarse o alejarse de determinados objetos. Asimismo, comienza a señalar los objetos alternando la mirada entre el objeto y el rostro del adulto. Finalmente, un logro muy importante de esta etapa es el de la atención compartida.

La *agencia intencional* es el modo de comprensión que corresponde a la segunda etapa, que se extiende desde el primer año hasta, aproximadamente, los tres años de edad. El infante empieza a atribuir estados mentales como deseos y sentimientos; es capaz de razonar de un modo no egocéntrico sobre los sentimientos y los deseos de las demás personas, y logra diferenciar entre las metas intencionales de una acción y sus consecuencias accidentales. Aún no puede representar estados mentales de forma independiente a la realidad física. La relación realidad y apariencia, mundo interno y mundo externo, todavía no se encuentra delimitada. Fonagy (2008) y Fonagy y Target (1997) sostienen que, en este periodo, el infante vive de manera simultánea dos formas de experimentar la relación entre la realidad externa y la realidad interna: el modo aparente (*pretend mode*) y el modo de equivalencia psíquica. En este último, la mente y el mundo es lo mismo, la relación entre ambos es especular, los estados mentales son reflejo del mundo y viceversa. No hay perspectivas. El modo aparente surge en el juego; sorprendentemente, en este caso el niño opera representacionalmente, puede creer que él es un superhéroe que va a salvar a una persona con su espada, la cual es, en realidad, un bastón. No obstante, se mantiene desconectado de la realidad; requiere separar la realidad interna de la realidad externa. La comprensión mentalística se logra cuando ambos modos se integran dando como resultado que los estados mentales sean representaciones que se relacionan con el mundo pero no de un modo equivalente. Así, en la siguiente etapa, el entendimiento se sofisticada e incluye la atribución de creencias y la consideración de perspectivas diferentes acerca del mundo.

En la etapa de la *agencia representacional*, entre los tres y los cinco años, el niño incluye a las creencias como estados mentales que motivan la acción. Asimismo, es capaz de diferenciar entre lo que percibe en un individuo y lo que podría estar sucediendo en la mente de ese mismo individuo. Logra atribuir falsas creencias a sí mismo y a los otros, y adquiere las capacidades necesarias para intercambios sociales nuevos como bromear, engañar y hacer trampa. Además, logra captar las emociones de los otros a través de la empatía. De otro lado, el niño empieza a preferir el juego con sus pares en lugar del juego con adultos. Es así como se va cerrando el periodo en el que la adquisición de las capacidades de mentalización se daba a partir de la agencia de una mente adulta y surge un nuevo periodo más largo en el que se promueve la comprensión de las mentes de otros con quienes se comparte similitudes en el sentido del humor y en los intereses.

Por último, la etapa de la *agencia autobiográfica*, que se presenta alrededor de los seis años de edad. En este momento hay grandes avances en el desarrollo de las habilidades para recordar y narrar experiencias y actividades intencionales de un modo organizado y coherente. El niño logra hacer un uso adecuado de la causalidad

y la temporalidad. Del mismo modo, la teoría de la mente se vuelve más sofisticada con la aparición de nuevas capacidades como las representaciones de segundo orden —comprender creencias erróneas sobre otras creencias—, la comunicación de sentimientos mezclados o encontrados —estar en conflicto—, modos de engaño más sutiles como las mentiras blancas, y la conciencia de que las expectativas de uno mismo pueden influir o sesgar la interpretación de un evento ambiguo. A partir de la elaboración de narrativas autobiográficas, los niños logran la comprensión de sí mismos y de los demás.

Existen dos funciones parentales esenciales que dan inicio a la secuencia de etapas antes mencionada: el reflejo (*mirroring*) y el *marking* (no hay traducción). El *marking* es la tendencia de los adultos a enfatizar con la voz y los gestos la emoción que el bebé transmite. Esta conducta le proporciona al infante la primera representación de su propia emoción, la cual usará como base para la identificación de sus emociones. Por su parte, el reflejo o *mirroring* se refiere a la transmisión del estado emocional del bebé a partir de la exageración de los gestos y las tonalidades de voz (Gergely, 2007).

Fonagy sostiene que el apego es fundamental para la adquisición de la teoría de la mente, y critica las posturas nativistas que plantean que el entorno social del infante dispara pero no determina su desarrollo. Si bien existe evidencia de una programación universal para el desarrollo de la teoría de la mente en un momento particular del desarrollo, la mayor parte de las investigaciones han encontrado que existe una variación transcultural con respecto al tiempo de su adquisición. Asimismo, factores como el modo de interacción entre los padres, la manera como estos hablan de las emociones a sus hijos, el espacio que le otorgan a las emociones en sus explicaciones sobre las acciones, y las creencias de los padres sobre la parentalidad, influyen en la teoría de la mente del niño (Fonagy, 2008).

Según Fonagy y Target (1997), las influencias sociales más importantes en el desarrollo de la teoría de la mente son el juego, el tipo de comunicación entre padres e hijos y el intercambio con los pares. El juego de simulación (*pretend*) favorece la comprensión de los estados mentales. Las representaciones que se originan durante dicho juego son, a la vez, diferentes de la realidad externa y compartida con otros individuos que son parte de la realidad: los compañeros de juego. Surge una conexión entre realidad y fantasía que empieza a perfilarse. El niño juega con la realidad y confía en que el otro con quien juega, que es real, puede compartir esa vivencia con él. De otro lado, el lenguaje tiene un rol central en la adquisición de la habilidad para mentalizar. Hablar de las intenciones y los sentimientos detrás de las acciones son prácticas cotidianas que se relacionan positivamente con la capacidad reflexiva del niño. Por último, el intercambio con hermanos y otros pares facilita la familiarización con perspectivas distintas a la propia, y permite la interacción con otras mentes.

No obstante, el señalamiento más importante para resaltar el papel del ambiente en la adquisición de la teoría de la mente se encuentra en los estudios que vinculan apego y mentalización. El punto principal es que la seguridad del apego entre el niño y sus cuidadores produce comprensión psicológica en la vida futura del niño.

Fonagy y Target (1997) plantean que los cuidadores *seguros* muestran una tendencia a *contener* los afectos de sus hijos al transformarlos de modo tal que sean más manejables y comprensibles para ellos. En contraste, el vínculo con un cuidador inseguro hace que el niño se identifique con un comportamiento defensivo, proveniente del cuidador, ya que si no lo hace perdería la cercanía con este. Cuando los padres son despreocupados tienden a desatender el afecto del niño, dado que les resulta demasiado movilizante y los hace recordar vivencias dolorosas. En el caso de los padres preocupados, la tendencia es a exagerar o tergiversar la experiencia emocional del niño. En cualquiera de los casos, no hay una sincronía entre la manera de decepcionar el afecto del padre y el modo de sentir del niño, lo cual da lugar a una sensación de falta de coherencia y de sentido del afecto al interior del sí mismo. Ello podría ocasionar en el niño una dificultad para representarse la experiencia afectiva (en el caso de cuidadores despreocupados) o la interiorización de construcciones extrañas sobre sí mismo (en el caso de cuidadores preocupados).

La adquisición de la teoría de la mente es parte de un proceso intersubjetivo entre el niño y sus cuidadores. La habilidad del niño para desarrollar una teoría de la mente depende de la observación de sus propios estados mentales en el mundo mental del adulto cuidador. Para que esto ocurra, el cuidador debe ofrecerse como alguien confiable que promueva la exploración mental del niño, y ser lo suficientemente sensible para captar y comunicar el sentido de los estados mentales en el niño durante las interacciones de rutina. Esta sensibilidad en los padres o cuidadores es un elemento esencial para el surgimiento de un vínculo de apego seguro en los hijos.

Cuando un cuidador es confiable, el niño siente la confianza suficiente para poder interpretar su conducta, ya que no encuentra nada de qué asustarse. En cambio, el niño con apego inseguro tiende a empobrecer sus capacidades reflexivas sobre los estados mentales de las otras personas, se aleja del intercambio intersubjetivo y, con ello, de la posibilidad de encontrar en el otro el medio para identificar sus propios estados mentales. El movimiento es defensivo, es decir, que el niño lo que está intentado hacer es protegerse. Así, prefiere refugiarse en la fantasía y la desconexión (características similares a las del cuidador) para poder defenderse y, a la vez, conservar el vínculo con sus cuidadores. El niño desorganizado o ambivalente desarrolla de manera prematura la capacidad para interpretar los estados mentales detrás de la conducta de sus padres en ciertas situaciones (particularmente las ansiógenas o violentas), pero no orienta la función reflexiva hacia sí mismo. Hay un desequilibrio

en el uso de las facultades mentales reflexivas ya que estas se concentran en los cuidadores, quienes, al ser percibidos como impredecibles, ambivalentes y hasta violentos, generan en el niño una hipervigilancia en relación a sus estados mentales. Adquiere algo como una teoría-teoría para comprender la conducta de los otros, pero no una capacidad para mentalizar propiamente dicha dado que el papel que esta tiene en la organización del *self* no se logra. El niño, en cambio, se protege ante la posibilidad de representarse a sí mismo en base a la imagen distorsionada (y a menudo maligna) que los padres o cuidadores le transmiten en las interacciones cotidianas.

Los niños que han sido expuestos a experiencias traumáticas suelen tener dificultades para delimitar adecuadamente entre realidad y fantasía. La sensación de agencia sobre la propia conducta se ve limitada, lo que afecta de manera sustancial la capacidad para diferenciar la propia experiencia de la ajena. Como resultado, algunas capacidades como la empatía se ven afectadas ya que el individuo no comprende la dimensión mental que subyace a la conducta de los otros, no produce una teoría del dolor ni la resonancia afectiva que esta tiene en su propia vivencia. Las personas con estructuras limítrofes de personalidad son un ejemplo para dar cuenta de una adquisición disfuncional de esta capacidad. La falta de sintonía de los padres ocasiona que sus estados mentales sean malinterpretados y reflejados por los mismos, y la persona los experimente como alienados. Intentará liberarse de ellos externalizándolos en sus vínculos con los demás. Las funciones prerreflexivas seguirán influyendo, de manera que la vivencia de la realidad quedará sujeta a las distorsiones de la fantasía interna y, de igual modo, la realidad interna será experimentada con la intensidad de la realidad externa.

La mentalización, entonces, abarca una serie de habilidades que permiten la autorreflexión y la comprensión de los demás en base a la atribución de estados mentales como los deseos, las creencias, los sentimientos, entre otros. Sin esta posibilidad de reflexionar sobre nosotros mismos y sobre los demás, como agentes intencionales, sería casi imposible entender el mundo social a nuestro alrededor.

2.4. Síntesis

La PED constituye un puente interesante entre los aspectos filogenéticos y los ontogenéticos de la cognición social. Busca analizar el desarrollo en la infancia, la niñez y la adolescencia desde una posición evolucionista por un lado y, por el otro, estudiar las conductas parentales que evolucionaron para proporcionar los cuidados necesarios durante esos periodos de desarrollo. Le interesa comprender cómo los mecanismos heredados —moldeados por la selección natural— se desarrollan y expresan a lo largo del ciclo vital de los individuos en interacción con su ambiente (Hernández & Bjorklund, 2008). Dentro de este modelo, Flinn y Ward (2005) consideran que

la competitividad social ha sido la presión de selección principal —desde que los seres humanos dominaran su medio ambiente—, que moldeó los procesos mentales centrales de los humanos, como la teoría de la mente y el lenguaje, ya que los más diferenciados cognitivamente, social y conductualmente tendrían mayor posibilidad de ejercer control sobre los demás individuos y, así, tener mayor acceso a los recursos.

Casi desde el inicio de la vida se observa una serie de logros íntimamente relacionados con la cognición social. Ya entre los dos y tres meses, los bebés entienden que sus actos producen una reacción en los demás y esperan que los adultos interactúen con ellos; van generando modelos de relación y son capaces de diferenciar entre las distintas expresiones emocionales de la madre. Asimismo, surgen los juegos cara a cara. Entre los nueve y doce meses, los infantes empiezan a entender la conducta de los otros como intencional y dirigida hacia una meta; va concretándose la capacidad para la atención conjunta, la cual es esencial para el desarrollo de la cognición social, y se manifiesta fundamentalmente con gestos protodeclarativos y gestos protoimperativos. Al año de edad, el uso de la referencia social es bastante evidente (Thompson, 2006; Carpendale & Lewis, 2010). A partir del año y medio comprenden que los deseos de los demás pueden ser diferentes de los de uno y que hay una relación entre el deseo y la emoción. A los dos años, los niños son capaces de representarse eventos y, así, tener expectativas sobre el futuro. De igual modo hacen referencias verbales sobre las emociones, sus causas y, también, los esfuerzos para su regulación. A partir de los tres años saben que las emociones se asocian a creencias y expectativas, y que las creencias pueden no ser consistentes con la realidad. Ya a los cinco años, los niños empiezan a percibir a los demás en términos de rasgos psicológicos y logran predecir el comportamiento futuro a partir de las características atribuidas.

Existen algunas teorías que se enfocan, sobre todo, en los procesos interpersonales que posibilitan el desarrollo de la cognición social. La teoría de la mentalización se ubica en esta categoría. La mentalización requiere del desarrollo de una serie de habilidades que permite al niño interpretar su conducta y la de los otros en términos de estados mentales intencionales —como los sentimientos, los deseos, las creencias, etcétera (Fonagy, 2008 y Busch, 2008)—. Para que ello se logre es necesario que el niño tenga un apego seguro, que le permitirá la exploración mental con un sentimiento de protección. Fundamentalmente, el cuidador deberá captar y comunicar el sentido de los estados mentales del niño durante los encuentros cotidianos puesto que el infante necesita hacer uso de las funciones ya maduras del cerebro adulto para organizar sus propios estados mentales. Es a partir de estos intercambios con otros que el individuo logra desarrollar las funciones mentales esenciales para un adecuado desarrollo y desenvolvimiento social (Siegel, 1999).