Teorema
Vol. XVI/2, 1997, pp. 35-49

Problemas del empirismo en la filosofía de la mente*

José Hierro S.– Pescador

RESUMEN

La experiencia de los estados mentales resulta central tan pronto como intentamos construir una ciencia de la mente. Los estados mentales parecen irreductibles a los estados físicos en la medida en que no son ni públicos ni computables. Desde un punto de vista epistemológico, los estados mentales tienen la peculiaridad de que de ellos no tenemos propiamente conocimiento, simplemente los tenemos. Desde el punto de vista de nuestra experiencia hay razones para rechazar la explicación fisicalista tanto como la explicación intencionalista, así como la reducibilidad a estados cerebrales. Hay dos formas de experiencia que son relevantes: la experiencia directa de los estados mentales propios y la experiencia indirecta de los estados mentales ajenos a través de la conducta y el habla de los demás.

ABSTRACT

Experience of mental states becomes central as soon as we attempt to construe a science of the mind. Mental states appear irreducible to physical states in as far as they are neither public nor computable. From an epistemological point of view mental states are peculiar in that we have no knowledge proper of them, we simply have them. From the point of view of our experience, there are reasons to reject the physicalist explanation as well as the intentionalist account and also to reject the reducibility of mental states to brain states. Two different forms of experience are relevant: direct experience of one's own mental states and indirect experience of other people's through their behaviour and speech.

Desde que Newton dio a la ciencia la forma en que la hemos recibido, la ciencia pareció explicar todo cuanto hay en la naturaleza convirtiéndose en el paradigma del conocimiento de lo material. Por ello, el mundo mental, que aparentemente es característico del ser humano, comenzó a resultar problemático. Ya Kant fue sensible a esta tensión cuando notó la diferencia entre concebir al hombre como un ser libre y concebirlo como sujeto a las leyes de la naturaleza, encomendando a la filosofía la tarea de mostrar "no sólo que ambos puntos de vista pueden coexistir, sino que ambos se deben pensar como necesariamente unidos". [Kant (1785), cap.3]

La distinción cartesiana entre sustancia pensante y sustancia extensa, que interaccionan en la glándula pineal, suministró el marco de lo que algu-

nos, como Ryle, han considerado una especie de doctrina oficial sobre el tema. No creo que Descartes hiciera otra cosa sino diseñar una teoría para explicar la relación entre dos clases de realidad –materia y espíritu– que venían siendo reconocidas por las sociedades humanas tanto en la vida cotidiana como en las creaciones culturales, incluyendo aquí la filosofía, la literatura y la religión. Según esta doctrina, que Ryle (1949) atribuía a Descartes, todo ser humano se compone de un cuerpo y una mente que sobrevive a la muerte del cuerpo. Lo mental se trataba como un mundo paralelo al mundo material y comparable con él. A esta forma de tratar la mente, Ryle llamó con ironía "mito del fantasma en la máquina". Es natural que con el nacimiento de la ciencia moderna no se resistiera la tentación de cosificar la mente. En el fondo, se trataba de objetivarla, a fin de hacer de ella objeto de la ciencia.

El empeño por justificar una ciencia de la mente no ha desaparecido. El significado paradigmático que la ciencia ha adquirido dentro de la teoría del conocimiento en la época moderna ha tenido esa consecuencia. Ello se ha puesto de manifiesto por lo pronto en el conductismo y en el fisicalismo, más recientemente en la ciencia cognitiva, y más recientemente aún en el intento de explicar la mente en términos de una mecánica cuántica convenientemente modificada, como ha hecho Penrose.

La aspiración a conseguir una ciencia de la mente ha ido unida al intento de vincular el estudio de la mente con la experiencia. Pero esta relación se puede manifestar al menos de dos maneras, sin duda conectadas entre sí: por un lado, en el conocimiento que tenemos de la mente; por otro lado, en el carácter de las pruebas que podemos ofrecer en apoyo de nuestras afirmaciones acerca de ella. En esta conexión está lo más peculiar de la mente. ¿Cómo conocemos la mente? Responder a esta pregunta obliga a distinguir. ¿Se trata de la mente propia o de la mente ajena? Si aceptamos que todos los seres humanos (al menos para empezar) tenemos mente, tendremos que reconocer que estamos hablando de algo que es común a todos, y por consiguiente tendremos que invocar alguna justificación que sea intersubjetiva. La dificultad se encuentra en que el conocimiento que tenemos de la mente propia tiene un carácter tan distinto del conocimiento que tenemos de la mente ajena que son irreductibles el uno al otro. De la mente propia tiene cada cual un conocimiento inmediato, directo, privado, e interno. No necesito de mis sentidos externos para conocer los estados de mi mente, esto es, mis estados psicológicos. Mis estados mentales los tengo, son el contenido de mi experiencia interna. Hay aquí una diferencia muy importante con respecto a la experiencia externa. En esta última, yo puedo señalar a su objeto, y cualquier otra persona cuyos sentidos externos funcionen correctamente lo percibirá. En el caso de la experiencia interna tal cosa no es posible. Puedo manifestar mi estado mental por medio de gestos o por medio de algún comportamiento coherente, o mejor aún por medio de palabras. Y de esta forma es como conocemos la mente de los demás. No tenemos experiencia

nocemos la mente de los demás. No tenemos experiencia directa de los estados mentales ajenos; solamente tenemos experiencia de sus manifestaciones externas. Y claro está que podemos equivocarnos y atribuir a alguien un estado psicológico que no tiene. Pero ¿podemos equivocarnos también respecto a nuestros propios estados mentales? Algunos han venido manteniendo que no, e incluso han tomado la incorregibilidad (en primera persona) como el carácter distintivo de lo mental. No a todos los autores, sin embargo, ha convencido esta tesis. Aquí habría que distinguir entre la imposibilidad de equivocarse, que habría que llamar infalibilidad, y la imposibilidad de mostrar a alguien que se ha equivocado, que sería propiamente incorregibilidad. Freud defendió que hay estados mentales (como ciertos deseos y ciertos sentimientos) que el sujeto niega tener, y que sin embargo tiene (por ejemplo en su subconsciente), y muchos que no comulgan con las teorías freudianas reconocen, no obstante, que con frecuencia nos engañamos respecto a nuestros verdaderos motivos y propósitos. Aparte de los argumentos de Armstrong (1968) y de J.C. Smart (1963), hay que recordar en el campo de los psicólogos las consideraciones críticas de Natsoulas (1970) así como los comentarios de Nisbett y Wilson (1977); estos últimos han mostrado que, para los procesos cognitivos superiores, los sujetos suelen carecer de pruebas introspectivas suficientes, las cuales sustituyen por prejuicios acerca de las supuestas causas de sus respuestas.

Pero sea o no sea infalible, ¿tiene cada persona un conocimiento privilegiado de sus estados mentales? No puedo decir "Ya sé que tengo un dolor", porque esto invita a la pregunta "¿Y antes no?" Mi afirmación sólo puede significar que antes yo no tenía un dolor. ¿Cómo puedo añadir a mi dolor un conocimiento? O tengo un dolor o no lo tengo. Wittgenstein escribió:

> ¿En qué sentido son mis sensaciones privadas? —Bien, sólo yo puedo saber si tengo realmente un dolor; otra persona sólo puede suponerlo.—En un sentido esto es erróneo, y en otro sentido es absurdo. Si estamos usando la palabra "saber" como se usa normalmente (¿y de qué otra manera la vamos a usar?), entonces otras personas muy a menudo saben cuándo tengo un dolor.—Sí, pero ¡no con la certeza con que —lo sé yo mismo!—No puede decirse de mí en absoluto (excepto tal vez como chiste) que yo *sé* que tengo un dolor. ¿Qué se supone que quiere decir esto—excepto tal vez que yo *tengo* un dolor?
> No se puede decir que otras personas conozcan mis sensaciones *sólo* por mi conducta–pues no se puede decir de mí que yo las conozca. Yo *las tengo*. Esto es lo correcto: tiene sentido decir de otras personas que dudan si yo tengo un dolor; pero no decirlo de mí mismo.[Wittgenstein (1953), § 246]

Ésta es la base de la infalibilidad tanto como de la incorregibilidad: si puedo dudar sobre los estados psicológicos de otra persona, pero no sobre los míos, entonces no tiene sentido que otra persona intente corregir mi autoads-

cripción de un estado psicológico negando que yo lo tengo. ¿Con qué argumento podría hacerlo? Solamente mostrando que mis palabras o mi comportamiento no lingüístico son incoherentes con el estado mental que me atribuyo. ¿Pero qué seguridad podemos tener respecto a la coherencia entre el comportamiento o las palabras de alguien y sus estados mentales? No damos a nuestras expresiones de carácter psicológico el significado que tienen en virtud de nuestros estados mentales que ellas designan.

Es absurdo decir que yo sé que tengo un dolor porque no tiene sentido decir que dudo si tengo un dolor, y es erróneo decir aquello porque tal afirmación manifiesta que estoy usando erróneamente el verbo "saber". Si a mí me parece que tengo un dolor, entonces lo tengo. Salvo en un caso: que yo esté usando el término "dolor" erróneamente, y lo que siento no sea lo que en mi comunidad lingüística se llama "dolor". ¿Cómo sé yo lo que designa esta palabra? No porque haya conectado la palabra con lo que siento, porque de esta manera nunca podría haber aprendido el significado de la palabra. Solamente puede haber una forma en que yo haya aprendido el significado del término "dolor": habiendo tenido experiencia de las manifestaciones externas a las que acompaña el uso de ese término [Wittgenstein (1953) § 244]. Y lo propio vale para todos los demás estados mentales. Es lo que resumió Wittgenstein en su lacónica afirmación:

> Un "proceso interno" requiere criterios externos. [Wittgenstein (1953) § 580]

Como los criterios para la adscripción de un estado mental nos dan su definición, quiere esto decir que los estados mentales se definen por sus manifestaciones externas, y en esto consiste el conductismo lógico de Wittgenstein. Se trata de un conductismo con estados mentales; hay estados mentales internos y manifestaciones externas, y estas últimas suministran los criterios definitorios de los primeros. Cualquier crítica a Wittgenstein que vaya más lejos de esto sería difícil de justificar, y esto es lo que muestra la excelente investigación del profesor García Suárez sobre este tema en su libro *La lógica de la experiencia* (1976).

Esta tesis implica que podemos tener estados mentales diversos sin saber cómo se designan, y por consiguiente sin haber adquirido un lenguaje. Es parte de esta tesis que los estados mentales poseen formas naturales y primitivas de expresión [Wittgenstein (1953)]. Si esto fuera así, la tesis parece correcta. ¿Pero lo es? ¿Cuál es la base de esa generalización? Supongamos un estado mental que carece de manifestaciones externas: no podríamos nombrarlo como no sea por analogía con otro que sí las tenga. Que alguna persona pueda tener un estado mental que no se manifiesta externamente es algo que únicamente la propia persona podrá decidir para sí misma si ocurre o no; porque si no hay manifestaciones externas, entonces nadie podrá llegar al conocimiento de tal estado mental excepto quien lo tiene, y esta persona

nocimiento de tal estado mental excepto quien lo tiene, y esta persona propiamente no lo conoce, simplemente lo tiene. De esta manera, la experiencia de los estados mentales resulta tener dos formas: la experiencia directa, que consiste en tenerlos, y es la experiencia que cada persona tiene de sus propios estados mentales; y la experiencia indirecta, que consiste en percibir sus manifestaciones externas, y es la que tenemos de los estados mentales ajenos, y la única que constituye propiamente conocimiento, porque admite la duda. Esta experiencia es la que nos suministra los criterios de alioadscripción de los estados mentales, y por tanto nos da su definición.

Que puede haber estados mentales concretos (instanciados) que carezcan de manifestaciones naturales o primitivas no hay razones para negarlo. Si los hay de hecho, tan sólo el sujeto que los tenga puede afirmarlo, pero no existiendo esas manifestaciones, tampoco el sujeto que los tiene puede nombrarlos ni describirlos, sino todo lo más describir a qué estados mentales manifestables se parece eso que ahora siente. Encontramos aquí una dificultad con la cual tropieza la teoría de la mente, y es un servicio que debemos agradecer a Wittgenstein el que haya llamado la atención sobre ello, pues contribuye a poner en cuestión el sentido y la eficacia de la llamada "introspección". La conclusión es que la introspección no puede ser otra cosa que la conciencia (el darse cuenta) de los estados mentales propios. Y el resultado de la introspección, así entendida, sólo puede ser comunicado o hecho público si se trata de estados mentales que poseen manifestaciones externas primitivas, porque únicamente en este caso podemos utilizar el lenguaje para hablar de nuestros estados mentales.

¿Y si no se trata de estados mentales concretos, de instancias de estados mentales, sino de tipos de estados mentales? La gran diferencia es que, en tal caso, tendremos algún término con el que referirnos a ellos, porque los tipos son, por definición, abstracciones, y no ocurren, no tienen lugar; lo que ocurre es una instancia. La pregunta anterior, por consiguiente, es la pregunta de si tenemos nombres de estados mentales cuyas instancias carezcan de manifestaciones externas primitivas. Si los tuviéramos, constituirían un contraejemplo a la tesis de Wittgenstein. Por ejemplo, ¿cuáles son las manifestaciones primitivas o naturales de la melancolía? Tal vez las haya: el gesto de la cara, la mirada, la forma desganada de hacer las cosas. ¿Pero no son estas también manifestaciones de la tristeza? ¿O es que tristeza y melancolía son el mismo tipo de estado mental? Podríamos tener palabras distintas para denotar el mismo tipo de estado mental sin habernos dado cuenta. Caer en la cuenta de ello puede ser otro servicio que debemos a Wittgenstein. Claro es que no habría por qué atribuir a los académicos de la lengua un talante wittgensteiniano, pero hay que notar que la melancolía está definida en el Diccionario de la Real Academia de la Lengua como "tristeza vaga, profunda, sosegada y permanente", la tristeza está definida como "calidad de triste", y triste a su

vez, como "afligido, apesadumbrado", y en su segunda acepción como "de carácter o genio melancólico". ¿Hemos aprendido el significado de las palabras "tristeza" y "melancolía" en conexión con las mismas manifestaciones externas?

La definición de los estados mentales no es la definición de las manifestaciones externas. ¿Cómo definirlos entonces? Consideramos diferentes tipos de manifestaciones externas o formas de conducta y diferentes tipos de acontecimiento que son causa indirecta de aquéllas. Entre la situación en que se encuentra el sujeto y su respuesta de comportamiento media su estado mental, que es efecto de la situación y causa directa del comportamiento. Como no se ha comprobado una relación clara entre tipos de estados mental y tipos de comportamiento, la definición de los estados mentales está aquejada de la imprecisión que hemos visto en el ejemplo de la melancolía. Hace ya muchos años, esta imprecisión fue recogida por Bloomfield, cuando escribió:

> Podemos definir el significado de una forma de habla con exactitud cuando este significado tiene que ver con algún tema del cual poseemos conocimiento científico. Podemos definir los nombres de los minerales, como cuando decimos que el significado ordinario de la palabra *sal* es "cloruro de sodio (NaCl)", y podemos definir los nombres de plantas o animales por medio de los nombres técnicos de la botánica o la zoología, pero no tenemos ninguna forma precisa de definir palabras como *amor* u *odio*, que se refieren a situaciones que no han sido clasificadas con exactitud. [Bloomfield (1933), p. 139]

Se manifestaba así la tesis de que la definición de las palabras depende del conocimiento científico. Esta posición, que ha tenido una reciente versión en los escritos de Kripke, era sin duda tributaria del fisicalismo neopositivista.

En Kripke, la tesis principal es que las definiciones recurren a propiedades esenciales y suministran, por ello, verdades necesarias. Si el dolor consiste, según afirman los neurólogos, en la estimulación de fibras nerviosas de tipo C, entonces esto es verdad necesariamente; es decir, la identidad entre el tipo de estado mental que llamamos "dolor" y el tipo de estado neurofisiológico que llamamos "estimulación de fibras nerviosas de tipo C" sería una identidad necesaria [Kripke (1972) p. 337]. Pero es claro que dolor es lo que siente el sujeto, su estado mental, y Kripke tiene que reconocer que no hay razón para mantener la necesidad de identificar el dolor con la estimulación de fibras C. Encontramos aquí, de nuevo, la dificultad que pudimos notar anteriormente, y que deriva del carácter subjetivo de los conceptos mentales.

Ya los positivistas del Círculo de Viena se enfrentaron a este problema cuando intentaron hacer una filosofía empirista de la mente. Así, Carnap (1932-33) defendió que la psicología no podría ser una ciencia a menos que tomara la física como modelo, cumpliendo así el programa fisicalista por el que trabajaban los miembros del Círculo de Viena. Para cumplir este programa, Carnap ponía como condición que las afirmaciones psicológicas fue-

ran traducibles a un lenguaje en el que los términos no lógicos tuvieran como referencia estados corporales, los cuales eran definidos en términos de disposiciones para reaccionar de cierta manera a estímulos determinados. De esta forma, el comportamiento seguía siendo la clave para la definición de los estados mentales, y la perspectiva conductista parecía resultar inevitable. De hecho, Carnap abandonó posteriormente, en [Carnap (1956)], esta explicación de los términos mentales como designativos de estados corporales caracterizados por disposiciones para la conducta, prefiriendo considerarlos como términos teóricos introducidos en la psicología como primitivos y conectados con los términos observacionales por medio de las necesarias reglas de correspondencia. En la base está implícita la idea de que hay una conexión lógica entre los conceptos mentales y los conceptos que se refieren al comportamiento.

Pero hay otra línea de origen muy anterior en la que se ha aspirado a fundamentar el estudio científico de la mente. En 1874 un pensador austríaco, Franz Brentano, publicó un libro titulado *La psicología desde un punto de vista empírico*, donde resucitaba el viejo concepto de intencionalidad, ya utilizado por pensadores medievales tanto árabes como cristianos, para construir sobre esa base la distinción entre el mundo físico y el mundo psicológico, y escribía así:

> Todo fenómeno psíquico se caracteriza por lo que los escolásticos de la Edad Media llamaban inexistencia intencional (o también mental) del objeto, y que nosotros, aunque no con expresiones del todo inequívocas, podríamos llamar relación a un contenido, dirección a un objeto (con lo cual no hay que entender que sea una realidad) o inmanente objetividad. [Brentano (1984), 1ª parte, libro 2, cap.I, secc. 5]

La intencionalidad ha hecho fortuna. Primero, suministró el marco para el desarrollo de la filosofía fenomenológica, y aunque ausente en la filosofía analítica durante muchos años (no hay una sola mención de este rasgo en *El concepto de lo mental* de Ryle ni tampoco en *The Analysis of Mind* de Russell) reaparece en los años cincuenta con fuerza inusitada, por ejemplo en Chisholm, para convertirse después en objeto primario de atención en autores como Searle, Dennett y Stalnaker. Ya Sellars, en 1956, en un artículo precisamente sobre el empirismo y la filosofía de la mente ("Empiricism and the Philosophy of Mind") imaginó un lenguaje que llamó *ryleano* porque su vocabulario descriptivo fundamental trata de propiedades públicas de objetos públicos localizados en el espacio y en el tiempo, y Sellars se preguntaba qué recursos deben añadirse a este lenguaje para que sus hablantes se puedan reconocer como animales que piensan, observan, y tienen sentimientos y sensaciones. Su respuesta es que lo primero que hay que añadir es el discurso

semántico, es decir, aquel en el que hablamos del significado de las palabras, y añadía:

> Pues es característico de los pensamientos su intencionalidad, referencia, o ser acerca de (*aboutness*), y está claro que el discurso semántico acerca del significado o la referencia de las expresiones verbales tiene la misma estructura que el discurso mentalista acerca de aquello de lo que tratan los pensamientos. Es por ello tanto más tentador suponer que la intencionalidad de los *pensamientos* puede encontrarse en la aplicación de las categorías semánticas a las actuaciones verbales públicas.[Sellars (1956), p. 201]

De esta forma, la atribución de estados mentales, como los diversos sentimientos, queda vinculada a la utilización de expresiones semánticas, tomándose este uso como equivalente a la intencionalidad. Pero aquí debe notarse que, aunque es cierto que el discurso semántico supone intencionalidad, e incluso tal vez sea su manifestación más clara, ésta es una intencionalidad de segundo grado, que corresponde a lo que los escolásticos llamaban "segundas intenciones", ya que la intencionalidad como tal no es una característica primariamente del lenguaje sino de la mente, y no podríamos dejar de atribuir intencionalidad a una persona porque careciera de lenguaje. En realidad, el problema de la intencionalidad es su exceso de capacidad explicativa.

El problema de Sellars es reconciliar la idea de que los pensamientos son episodios *internos* que no requieren ni manifestaciones de conducta ni imágenes lingüísticas, y a los cuales nos referimos con el vocabulario de la intencionalidad, con la idea de que las categorías de la intencionalidad son, en el fondo, categorías semánticas de carácter lingüístico. Su sugerencia es que los pensamientos son introducidos como episodios teóricos que atribuimos a otros sobre la base de su comportamiento, y que cada cual atribuye a sí mismo sobre la base del comportamiento propio. El comportamiento resulta así la prueba de tales episodios teóricos, pero no los excluye.

¿En qué casos debemos considerar el comportamiento de un organismo como prueba de estados intencionales y en cuáles no? ¿Debemos atribuir a la araña la creencia de que por medio de su tela obtendrá los insectos que le sirvan de alimento? ¿Nos sirven los conceptos teóricos de estados intencionales para predecir la conducta de la araña? Si nos sirven, entonces deberemos adoptar frente a ella lo que Daniel Dennett ha llamado "actitud intencional" (1987). Esta actitud consiste en tratar el objeto cuya conducta se quiere predecir como un agente racional, que tiene creencias, deseos y demás estados mentales intencionales. La tesis de Dennett es que cualquier objeto cuyo comportamiento sea predicho por medio de tal estrategia es un sujeto de creencias (*believer*), y por tanto un sistema intencional.

Es cierto que de hecho consideramos a las personas incluidas en esta categoría, y con respecto a ellas, los conceptos intencionales son aquellos conceptos teóricos que constituyen el marco de una teoría de la persona en-

tendida como sistema intencional. ¿Es ésta la única teoría aplicable a las personas? No. Pero parece la mejor a la vista de las alternativas posibles. Predecir el comportamiento de alguien con razones astrológicas no ofrece garantías suficientes, se refuta fácilmente y con frecuencia, y además resulta del todo incongruente con el marco cultural en el que nos movemos, dentro del cual la ciencia moderna tiene un lugar de privilegio. Predecirlo sobre la base de la constitución físico-química de las personas tampoco parece una estrategia de éxito, al menos en nuestro actual nivel de conocimientos. Cabría recurrir a la estrategia que Dennett llama "del diseño", y con la cual predecimos el comportamiento de los objetos presuponiendo que éstos se comportarán tal y como han sido diseñados para comportarse. Esta teoría es particularmente apta para toda clase de artefactos, aunque Dennett no tiene inconveniente en extenderla a objetos biológicos como los animales y las plantas.

Con ello tenemos una primera respuesta a la pregunta anterior. Podemos predecir que la araña producirá su tela sobre la rama del árbol porque la evolución natural ha diseñado a este animal de manera que éste pueda obtener su alimento capturando otros animales por ese medio. Tenemos suficiente con esta teoría del diseño y no necesitamos atribuir a la araña creencias de ninguna clase. ¿Cambiaría en algo nuestra explicación si le atribuyéramos creencias? Nuestra predicción podría ser más exacta y detallada, pero difícilmente se cumpliría mejor. La principal dificultad estaría en saber qué creencias y qué deseos atribuir a cada araña en cada situación. La falta de coincidencias biológicas excluye la existencia de una comunidad entre las arañas y las personas que sea suficiente para predecir su comportamiento en términos de deseos y creencias.

La consecuencia es que, por definición, la estrategia intencional solamente tiene sentido aplicarla a las personas. ¿Por qué? Porque es una estrategia pensada para objetos que tienen creencias y deseos, y únicamente conocemos una categoría de tales objetos, a saber, las personas. Y esto es así porque hemos encontrado fácil explicar y predecir el comportamiento humano en términos de creencias y deseos, y no nos sentimos tentados a aplicar estas categorías a otros objetos a menos que estén muy próximos a nosotros, como ocurre con los mamíferos superiores y los robots. Los primeros porque se hallan cerca en la escala biológica; los segundos porque los hemos fabricado nosotros. Por lo que respecta a los robots, la estrategia adecuada parecería ser la del diseño, pero nuestro propio orgullo nos hace acariciar la idea de haber sido capaces de crear la mente, y nos hemos enzarzado en una compleja e interminable discusión sobre si los robots piensan o no piensan, e incluso sobre si tienen o no sentimientos. En el fondo, la cuestión es cómo vamos a definir estos conceptos y a qué tipo de objetos estamos dispuestos a aplicarlos. Puesta de otra manera, la cuestión es si nos queremos colocar más cerca de los perros o de las máquinas.

Algunos autores han considerado que la intencionalidad humana es original mientras que la del robot es una intencionalidad derivada. Esta es una distinción que resulta obligado hacer si se aplica la categoría de intencionalidad de manera unitaria a personas y a robots. Hay que notar, sin embargo, que, en la teoría de Dennett, la intencionalidad de las personas no es original sino derivada. Derivada ¿de dónde? De la evolución natural. Sólo a ésta le corresponde una intencionalidad propiamente original. Es una tesis que muestra tanto la excesiva amplitud del concepto de intencionalidad como el peligro que encierra su aplicación. Porque si estamos tomando la intencionalidad como la característica definitoria de los estados mentales, entonces tenemos que considerar la evolución natural como una entidad unitaria, un objeto propiamente dicho, al cual atribuir estados mentales de los cuales las personas reciban la intencionalidad que poseen. Y naturalmente esto tan sólo puede aceptarse como una metáfora. No somos objetos creados por la naturaleza de la forma en que las máquinas han sido creadas por los hombres. La naturaleza no es una entidad, sino un conjunto de entidades en proceso de evolución, y nosotros somos parte de ese conjunto en proceso. No hay todas las cosas naturales y además la naturaleza, sino cosas u objetos naturales en proceso de evolución, y algunos de esos objetos resultan tener esa propiedad que hemos venido en llamar "intencionalidad", y la tienen como resultado de la propia evolución natural, igual que, también como resultado de la evolución natural, esos objetos son animales que caminan erguidos sobre dos de sus miembros, objetos que acaso pueden ser descritos como artefactos diseñados por la naturaleza para asegurar la supervivencia de los egoístas genes, como dice Dawkins (1976). La entificación de la naturaleza es sugerida de modo implícito por Dennett cuando, considerando las razones que presiden la evolución natural, escribe: "La naturaleza ha apreciado estas razones sin representárselas." [Dennett (1987), p. 317]. Pero a renglón seguido se corrige para reconocer que no se trata de un objeto sino de un proceso, añadiendo: "Y el propio proceso del diseño es la fuente de nuestra intencionalidad propia", con lo cual reconoce que la intencionalidad brota del diseño evolutivo de la naturaleza. En mi opinión, esto implica aceptar que la estrategia del diseño, aplicada al devenir natural, es más básica que la estrategia intencional. O dicho de otra forma: que haya objetos con mente, y por tanto con estados intencionales, es un resultado de la evolución natural, y no hay razón para asombrarse de ello más de lo que la hay para asombrarse de que haya objetos con un cerebro que tenga el tamaño, peso y complejidad que tiene el cerebro humano.

Es una tesis ampliamente defendida que los estados mentales son estados del cerebro, y esta tesis nos obligaría a aceptar que, cuando sentimos tristeza, tenemos experiencia de un cierto estado del cerebro, en verdad distinto de aquel estado del cerebro que experimentamos cuando sentimos alegría.

También tendríamos que aceptar que pensar en la deducción trascendental es experimentar un cierto estado de las neuronas. Son afirmaciones tan extrañas a la psicología popular que producen un espontáneo movimiento de rechazo en cualquiera que no se haya comprometido previamente con la teoría de la identidad entre el cerebro y la mente. La razón de ese rechazo está en que la experiencia de los estados mentales y la experiencia de los estados cerebrales son a primera vista irreductibles entre sí. Explicamos los estados cerebrales con la misma perspectiva que usamos para explicar otros estados del organismo: determinadas sustancias químicas que interactúan entre sí transmitiendo impulsos eléctricos. Y ésta es una descripción que presupone un punto de vista externo, porque se trata de una situación percibida por los sentidos externos.

De aquí procede la dificultad en identificar el dolor con la estimulación de las fibras nerviosas de tipo C. Según la concepción de Kripke [1972, p. 337], si esta identidad es verdadera, entonces es necesaria, y ello implica que no puede haber una estimulación de fibras C que no sea un dolor, ni tampoco un dolor que no sea una estimulación de fibras C. Pero esta conclusión es contraintuitiva, puesto que la sensación de dolor podría ser producida por la estimulación de otros receptores nerviosos distintos de las fibras C sin dejar de ser dolor, ya que el término "dolor" nombra una sensación (estado interno) y no una cierta configuración de estímulos y zonas del organismo (estado externo). Es la diferencia que hay para Kripke [(1972), pp. 338 y ss.] entre el dolor y el calor, pues tomando el término "calor" como nombre de un fenómeno externo y no de una sensación, no hay dificultad en afirmar la identidad entre el calor y el movimiento de las moléculas como identidad necesaria. En la presencia de calor se puede o no sentir calor según el estado del sujeto, pero en presencia del dolor propio no puede dejar de sentirse dolor, porque el término "dolor" nombra precisamente lo que el sujeto siente. Estas consideraciones pueden generalizarse a todos los estados mentales, puesto que todos ellos consisten en experiencias del sujeto. Y esto es lo peculiar del mundo mental: que la experiencia no es un medio para acceder a los estados mentales, sino que es los propios estados mentales. En una vena muy parecida al párrafo de Wittgenstein que cito en otro lugar, escribe Kripke:

> Estar en la misma situación epistémica que habría si uno tuviera un dolor es tener un dolor [Kripke (1972) p. 339].

La argumentación de Kripke va dirigida contra la tesis de la identidad entre estados mentales y estados cerebrales, mostrando que tal identidad tendría que ser necesaria, pero que, por el contrario, en la correspondencia entre un estado del cerebro y un estado mental "parece haber un cierto elemento obvio de contingencia" [Kripke (1972) p. 341].

En resumen, conocemos los estados mentales propios en forma directa en cuanto que los tenemos, y los ajenos por medio de nuestros sentidos externos a través de las manifestaciones de aquellos. Conocemos el cerebro propio por los estados mentales que produce, pero conocemos el cerebro ajeno por sus manifestaciones en la conducta del sujeto. El comportamiento es manifestación de los estados mentales así como de los estados cerebrales que producen a estos últimos, y aun ocurriendo así tanto en el caso propio como en el caso ajeno, la diferencia entre ambos casos impone la siguiente diferencia epistemológica que es característica del mundo de la mente: para acceder a los estados mentales ajenos, y por tanto a los estados cerebrales del sujeto en cuestión, tenemos que recurrir a una vía indirecta que pasa por su comportamiento, mientras que en el caso propio hay una vía directa a los estados mentales, que consiste en tenerlos, y a través de ellos una vía indirecta a los estados cerebrales que son causa de aquéllos. En el caso propio nunca se requiere recurrir a la conducta ni utilizar los sentidos externos.

Considerando el tema de la conciencia, algunos han comparado la relación entre el cerebro y la conciencia con la relación entre el estado líquido del agua y su composición química como H_2O. La idea parece ser que la conciencia es al cerebro como la liquidez es al agua, a saber, el estado natural en que se encuentra. Puede que la comparación valga para la conciencia, que es el estado normal del cerebro en situación de vigilia, y que acompaña a los estados mentales típicos. Pero a la hora de caracterizar cada tipo de estado mental, pienso que tendríamos que considerarlo como un tipo de estado cerebral que se manifiesta en la experiencia interna de una manera determinada (como tristeza, como alegría, como amor, como odio) y que tiene las consecuentes manifestaciones en la experiencia externa (gestos, mirada, palabras, tono de voz, forma de andar).

Aparte de su manifestación en la experiencia interna, que es su carácter subjetivo, tal vez la característica más llamativa de la mente sea el carácter no computable de los procesos mentales, que para algunos haría imposible un tratamiento científico de la mente. Es el tema que ha venido tratando Roger Penrose en sus últimos libros, de 1989 (*The Emperor's New Mind*) y de 1994 (*Shadows of the Mind*). Penrose trata en particular de la conciencia, y podemos tomarla como paradigma de la mente humana. Hay que distinguir en la mente dos aspectos: el contenido y la conciencia. El contenido, que equivale a la intencionalidad, es para Dennett más básico que la conciencia en la construcción de la teoría de la mente, pero esta posición me parece dudosa puesto que la conciencia acompaña a todos los estados intencionales, exceptuando claro está los estados subconscientes, y acompaña también a estados que no son intencionales como el dolor.

Es notable que, después de haberse dedicado la mayor parte de la atención a la intencionalidad, los libros sobre la conciencia se hayan multiplicado

en los últimos años. Parece que la mente no puede caracterizarse como un proceso computable, al que se le pueda aplicar un algoritmo o procedimiento mecánico de decisión. Hay en los procesos mentales un elemento que conduce a resultados aleatorios como los descubrimientos y los inventos, un elemento que está presente de modo más llamativo en la creación artística, y es un elemento que explica la libertad de elección en la conducta. Esto está en contradicción con la ciencia clásica, pero Penrose no duda en introducir en la mecánica cuántica las modificaciones necesarias para asegurarse de tener una explicación científica que explique cómo el cerebro, y en particular los microtúbulos del citoesqueleto de las neuronas, pueden originar la conciencia.

¿Qué pruebas hay de que sea esa concreta porción del cerebro la que está asociada a la conciencia? Para Penrose la prueba está en que se produce la pérdida de conciencia suministrando al organismo ciertas sustancias químicas que aparentemente actúan sobre los microtúbulos del citoesqueleto de las neuronas: son las anestesias, como el éter y el cloroformo. Penrose ha recordado también que estas sustancias inmovilizan a los animales unicelulares como la ameba y el paramecio. Y cabe entonces preguntarse: puesto que toda célula (eucariótica) posee citoesqueleto (a diferencia de las células procarióticas como las bacterias y las algas verdiazules), ¿cooperan todas las células de nuestro cuerpo a las funciones mentales? ¿Tienen mente los organismos que carecen de neuronas pero tiene otras células? ¿Tienen mente los animales unicelulares como la ameba y el paramecio? La respuesta a estas últimas preguntas parece que debe ser positiva si aceptamos como manifestación de la mente ciertos movimientos y funciones biológicas de la célula en cuestión. Pero como las manifestaciones de la mente humana son tan complejas, Penrose no tiene inconveniente en admitir que, para la mente humana, no sólo cuentan los microtúbulos del citoesqueleto de las neuronas, sino también la compleja organización de estas últimas. Pero sobre este punto carecemos de un conocimiento más preciso.

Y nótese otra cosa más: ¿significa atribuir mente lo mismo que atribuir conciencia? Aquí se impone distinguir la conciencia, que es la autopercepción y el darse cuenta de las respuestas a los estímulos, y que pensamos que es característica definitoria de los seres humanos, de otras manifestaciones de la mente que podríamos atribuir sin esfuerzo a otros animales, como es la resolución de problemas o la identificación de depredadores.

Pero hay más aún: tal vez sólo ciertas formas de conciencia son privativas y características de los humanos. Tenemos que empezar por distinguir diferentes formas de la conciencia. Para empezar, la conciencia fenoménica (*phenomenal consciousness*) y la conciencia cognitiva (*cognitive consciousness*). Esta última incluye la conciencia del yo (*self-consciousness*), la conciencia sobre los estados mentales propios o conciencia reflexiva (*reflective consciousness*) y la conciencia de acceso (*access consciousness*), la cual tie-

ne como objeto el control del comportamiento y del habla. No me parece claro con qué criterios podamos atribuir o negar conciencia cognitiva a un animal. En cambio, no veo razones para negar conciencia fenoménica a otros animales, ya que la conciencia fenoménica incluye el modo en que se ven, se oyen y se huelen las cosas, y también el modo en que se siente el dolor [véase Ned Block (1993)].

No siempre se hacen estas distinciones, y es la razón por la que Ned Block critica la teoría de la conciencia de Dennett (1991). En primer lugar, porque el modelo de borradores múltiples (*multiple drafts model*) mantiene que hay representaciones paralelas distintas que tienen acceso al razonamiento, a la verbalización y a la conducta, y en esta medida hay que tomarlo como un modelo de la conciencia de acceso, aunque Dennett suele hablar de la conciencia como si la estuviera tomando en el sentido fenoménico. En segundo lugar, porque Dennett defiende al mismo tiempo que la conciencia es un complejo de memes, esto es, de unidades de transmisión cultural, paralelas a los genes, según la terminología introducida por Richard Dawkins (1976), y esto parece más bien una concepción del contenido de la mente que de la conciencia, si se acepta la distinción entre contenido y conciencia introducida desde un principio por el propio Dennett en la base de la teoría de la mente.

La esencial conexión entre la mente y la conciencia, unida a la tesis de que la conciencia es esencialmente subjetiva, ha conducido a Searle (1992) a la conclusión de que la ontología de lo mental es una ontología de la primera persona, y esta conclusión va unida a la asimetría epistemológica existente entre la experiencia que cada cual tiene de los estados mentales propios y la que tiene de los estados mentales ajenos. Este dualismo epistemológico es característico de la filosofía de la mente y obedece a las diferencias que hay que reconocer entre la experiencia que consiste en tener estados mentales y la experiencia que consiste en percibir acontecimientos externos. Seríamos infieles a nuestra experiencia si, a la hora de hacer una descripción completa del mundo, nos limitáramos a describir estados físicos. El problema es que no parece que podamos describir nuestros estados mentales sin recurrir a sus manifestaciones externas, porque sólo en conexión con éstas damos significado a las expresiones por medio de las cuales hablamos de aquéllos. Tal vez estamos condenados a ser infieles a nuestra experiencia.

Departamento de Lingüística, Lenguas Modernas,
Lógica y Filosofía de la Ciencia
Universidad Autónoma de Madrid
Ciudad Universitaria de Cantoblanco
28049 Madrid

NOTAS

*Versión corregida de la ponencia leída en el congreso sobre "Problemas del empirismo" celebrado en la Universidad de Oviedo en abril de 1996. Mi agradecimiento a los organizadores del congreso por su amable invitación.

REFERENCIAS BILBIOGRÁFICAS

ARMSTRONG, D. (1968), *A Materialist Theory of the Mind*, Londres, Routledge.
BLOCK, N. (1993), "Review of Daniel Dennett's *Consciousness Explained*", *The Journal of Philosophy*.
BLOOMFIELD, L. (1933), *Language*, Nueva Yorh, Holt, Rinehart & Winston.
BRENTANO, F. (1874), *Psychologie vom empirischen Standpunkt,* Leipzig.
CARNAP, R. (1932), "Psychologie in physikalischer Sprache", *Erkenntnis*, vol.3.
——(1956), The Methodological Character of Theoretical Concepts", en: H. Feigl *et al.*, (eds.), *Minnesota Studies in the Philosophy of Science*, vol. I, Minneapolis, Univ. of Minnesota Press pp. 38-76.
DAWKINS, R. (1976), *The Selfish Gene*, Oxford University Press
DENNETT, D. (1987), *The Intentional Stance*, Cambridge, Mass. MIT Press,
GARCÍA SUÁREZ, A. (1976), *La lógica de la experiencia,* Madrid Tecnos,
KANT, I. (1785), *Grundlegung zur Metaphysik der Sitten.*
KRIPKE, S. (1972), "Naming and Necessity", en Davidson, D. y Harman, G (eds), *Semantics of Natural Language*, Dordrecht, Reidel.
NATSOULAS, T. (1970), "Concerning Introspective Knowledge", *Psychological Bulletin*.
NISBETT, R. AND WILSON,T., (1977), "Telling More Than We Can Know: Verbal Reports on Mental Processes", *Psychological Review.*
PENROSE, R. (1989), *The Emperor's New Mind*, Oxford University Press.
—— (1994), *Shadows of the Mind*, Oxford University Press.
RUSSELL, B. (1921), *The Analysis of Mind*, Londres, Allen & Unwin.
RYLE, G. (1949), *The Concept of Mind*, Londres Hutchinson.
SEARLE, J. (1983), *Intentionality*, Cambridge University Press.
—— (1992), *The Rediscovery of the Mind*, Cambridge, Mass, MIT Press.
SELLARS, W. (1956) "Empiricism and the Philosophy of Mind", en H. Feigl y M. Scriven (eds.), *Minnesota Studies in the Philosophy of Science*, vol. I, Minneapolis, Univ. of Minnesota Press.
SMART, J. C. (1963), *Philosophy and Scientific Realism*, Londres, Routledge.
WITTGENSTEIN, L. (1953), *Philosophische Untersuchungen*, Oxford, Blackwell.