

## 1. CEREBROS EN UNA CUBETA

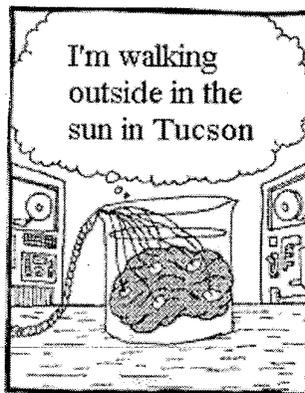
Una hormiga se arrastra lentamente sobre la arena. Conforme avanza va trazando en ésta una línea. Por puro azar, la línea se desvía y vuelve sobre sí misma, de tal forma que acaba pareciendo una reconocible caricatura de Winston Churchill. ¿Ha trazado la hormiga un retrato de Winston Churchill, un dibujo que *representa* a Churchill?

La mayoría de la gente, tras reflexionar un poco, contestaría que no. Después de todo, la hormiga nunca ha visto a Churchill, ni siquiera un retrato suyo, ni tampoco tenía intención de representarlo. Simplemente trazó una línea (y ni siquiera este acto fue intencional), línea que nosotros podemos ver como un retrato de Winston Churchill.

Podemos expresar esto afirmando que la línea no representa<sup>1</sup> por sí misma. La semejanza (de una especie muy complicada) con las facciones de Winston Churchill no es condición suficiente para que algo represente o se refiera a Churchill. Tampoco es condición necesaria: en nuestra comunidad, la forma impresa «Winston Churchill», las palabras «Winston Churchill», en tanto que pronunciadas, y muchas otras cosas, se usan para representar a Churchill (aunque no pictóricamente), si bien no tienen el tipo de semejanza con Churchill que sí tiene un retrato —o incluso un dibujo esquemático. Si la semejanza no es condición necesaria ni suficiente para que alguna cosa represente a otra, ¿cómo demonios puede una cosa representar (o estar en un lugar de, etc.) otra diferente?

La respuesta puede parecer fácil. Supongamos que la hormiga ha visto a Winston Churchill, y supongamos que tiene la inteligencia y la habilidad suficientes para dibujar un retrato suyo. Supongamos que ha elaborado la caricatura *intencionalmente*. Entonces la línea habría representado a Churchill.

Por otra parte, supongamos que la línea tiene la forma WINSTON



<sup>1</sup> En este libro, los términos «representación» y «referencia» aludirán siempre a la relación que se da entre una palabra (u otra clase de signo, símbolo o representación) y algo que existe efectivamente (esto es, no precisamente un «objeto del pensamiento»). Hay un sentido de «refiere» según el cual puedo «referirme» incluso a lo que no existe, mas no lo emplearé aquí. Una palabra más antigua para denominar lo que yo llamo «representación» o «referencia» es *denotación*. En segundo lugar, seguiré la costumbre de los lógicos modernos y usaré «existe» para significar «existe en el pasado, presente o futuro». Así pues, Winston Churchill «existe» y podemos «referirnos» o «representar» a Winston Churchill, aun cuando ya no esté vivo.

CHURCHILL, y que este hecho es un mero accidente (pasando por alto que es bastante improbable). Entonces los «caracteres impresos» WINSTON CHURCHILL no habrían representado a Winston Churchill, a pesar de que sí lo hacen cuando aparecen hoy en casi todos los libros.

De forma que puede antojársenos que lo que se necesita para la representación, o lo que se necesita principalmente para la representación, es la *intención*.

Pero para tener la intención de que *algo*, siquiera el lenguaje privado (incluso las palabras «Winston Churchill» repetidas mentalmente y no oídas) *represente* a Churchill, debo ser capaz de *pensar en* Churchill, para empezar. Si las líneas en la arena, los ruidos, etc., no pueden representar nada «en sí mismos», entonces ¿cómo es que pueden hacerlo las formas del pensamiento? ¿Cómo puede el pensamiento alcanzar y «aprehender» lo que es externo?

Algunos filósofos han dado un salto desde estas reflexiones hasta lo que ellos consideran como una prueba de *la naturaleza esencialmente no-física* de la mente. El argumento es simple; lo que dijimos acerca de la curva de la hormiga también se aplica a cualquier objeto físico. Ningún objeto físico tiene por sí mismo la capacidad de referirse a una cosa más bien que a otra; no obstante, es obvio que los *pensamientos de la mente* sí lo logran. De modo que los pensamientos (y por ende, la mente) poseen una naturaleza esencialmente distinta de la de los objetos físicos. Tienen la característica distintiva de la *intencionalidad* —pueden referirse a otras cosas; ningún objeto físico tiene «intencionalidad», salvo la intencionalidad que se deriva de su uso por parte de una mente. O eso se pretende. Pero esto es ir demasiado deprisa; postular misteriosos poderes mentales no resuelve nada. A pesar de todo el problema es real. ¿Cómo es posible la intencionalidad? ¿Cómo es posible la referencia?

#### TEORIAS MAGICAS DE LA REFERENCIA

Hemos visto que el «dibujo» trazado por la hormiga no tiene conexión necesaria con Winston Churchill. El mero hecho de que el dibujo mantenga cierta « semejanza » con Churchill no lo convierte ni en un retrato real ni en una representación de Churchill. Salvo que la hormiga sea una hormiga inteligente (y no es el caso) y sepa algo con respecto a Churchill (y tampoco es el caso), la curva que trazó no es un dibujo, ni tan siquiera una representación de algo. Ciertos pueblos primitivos creen que algunas representaciones (en particular los *nombres*) tienen una conexión necesaria con sus portadores; creen que saber el «verdadero nombre» de alguien o algo les otorga poder sobre ese alguien o algo. Este poder procede de una *conexión mágica*

entre el nombre y el portador del nombre; pero una vez que nos percatamos de que el nombre *sólo* tiene una conexión contextual, contingente y convencional con su portador, es difícil ver por qué el conocimiento del nombre ha de tener alguna significación mística.

Es importante darse cuenta de que a las *imágenes* mentales, y, en general, a las representaciones mentales, les ocurre lo mismo que a los dibujos físicos; la conexión que tienen las representaciones mentales con lo que representan no es más necesaria que la que tienen las representaciones físicas. La suposición contraria es un vestigio del pensamiento mágico.

El problema quizá se capte más fácilmente en el caso de las imágenes mentales (quizá el primer filósofo que captó la enorme importancia de este problema, pese a no ser realmente el primero en barajarlo, fue Wittgenstein). Supongamos que en alguna parte existe un planeta en el cual se han desarrollado seres humanos (o han sido depositados allí por extraños cosmonautas). Supongamos que esos humanos, si bien son como nosotros, nunca han visto un *árbol*. Supongamos que nunca se han imaginado un árbol (la única vida vegetal que existe en su planeta son los líquenes). Supongamos que cierto día, una nave que pasa por su planeta sin establecer contacto con ellos, arroja sobre éste el dibujo de un árbol. Imaginémosles devanándose los sesos ante el dibujo. ¿Qué demonios es esto? Se les ocurre toda clase de especulaciones: un edificio, un baldaquín, e incluso alguna especie de animal. Pero supongamos que ni siquiera se aproximan a saber de qué se trata.

Para *nosotros* la pintura es la representación de un árbol. Para aquellos humanos el dibujo únicamente representa un objeto extraño, de naturaleza y función desconocidas. Supongamos que, como resultado de ver el dibujo, uno de ellos tiene una imagen mental que es exactamente como mis imágenes mentales de los árboles. Su imagen mental no es la *representación de un árbol*. Sólo es la representación del extraño objeto (el que sea) que representa la misteriosa pintura.

Pese a esto, alguien podría argumentar que la imagen mental es *de hecho* la representación de un árbol, ya que, en primer lugar, el dibujo que provocó tal imagen mental era la representación de un árbol. Hay una cadena causal desde los árboles reales hasta la imagen mental, aun cuando esta sea muy extraña.

Sin embargo, podemos imaginarnos la ausencia de esta cadena. Supongamos que el «dibujo de un árbol» que la nave espacial arrojó no era en realidad el dibujo de un árbol, sino el resultado accidental del derrame de algunas pinturas. Aun cuando fuese exactamente igual al dibujo de un árbol, en realidad no sería el dibujo de un árbol en un grado mayor que la «caricatura» de la hormiga era un retrato de Churchill.

Podemos incluso imaginar que la nave espacial que arrojó el «dibujo» procedía de un planeta en el que no se sabía nada sobre los árboles. En tal caso, pese a que esos humanos tendrían imágenes cualitativamente idénticas a mi imagen de árbol, esas imágenes no representarían más a un árbol que a cualquier otra cosa arbitraria.

Lo mismo ocurre con las *palabras*. Un discurso impreso podría parecer una descripción perfecta de un árbol, pero si fueron los monos quienes lo produjeron golpeando fortuitamente las teclas de una máquina de escribir durante millones de años, entonces las palabras de ese discurso no se refieren a nada. Si alguien las memorizase y las repitiera mentalmente sin entenderlas, entonces cuando fuesen pensadas tampoco se referirían a nada.

Imaginemos que la persona que está repitiendo mentalmente estas palabras ha sido hipnotizada. Supongamos que tales palabras están en japonés, y que al hipnotizado se le ha dicho que entiende ese idioma. Supongamos que cuando piensa esas palabras experimenta algo así como un «sentimiento de comprensión». (Aunque si alguien irrumpiese en su flujo mental y le preguntase *qué significan* las palabras que está pensando, descubriría que no podría decirlo.) Quizá la ilusión sea tan perfecta que incluso podrá engañar a un telépata japonés. Pero si no es capaz de emplear las palabras en los contextos apropiados ni de responder preguntas con respecto a lo que «piensa», entonces no las entendió.

Combinando estas historias de ciencia-ficción, podemos idear un caso en el que alguien piensa ciertas palabras que constituyen de hecho la descripción de un árbol en algún lenguaje, y simultáneamente tiene unas imágenes mentales apropiadas, pero *ni* comprende las palabras *ni* sabe lo que es un árbol. Incluso podemos imaginar que las imágenes mentales fueron provocadas por un derrame de pintura (aunque la persona ha sido hipnotizada e inducida a pensar que son imágenes de algo apropiado a su pensamiento —sólo que, si se le pregunta de qué son imágenes, no podría responder). Y podemos imaginar que ni el hipnotizador ni el hipnotizado han oído hablar del lenguaje en el que este último está pensando— quizá sea una mera coincidencia el que estas «oraciones sin sentido», tal y como las considera el hipnotizador, sean la descripción de un árbol en idioma japonés. En resumidas cuentas, cualquier cosa que pase ante su mente puede ser cualitativamente idéntica a lo que estaba pasando por la mente de un hablante japonés que pensaba *verdaderamente* en árboles— pero ninguna de ellas se referiría a árboles.

Todo esto es realmente imposible, por supuesto, del mismo modo que es realmente imposible que los monos mecanografién por casualidad una copia de *Hamlet*. Y esto es afirmar que las posibilidades en contra son demasiado altas como para que este suceso realmente ocurra. A pesar de todo, no es lógicamente imposible, ni siquiera físi-

camente imposible. *Podría* suceder (es compatible con las leyes de la física, y quizá también con las condiciones actuales del universo, si hubiese seres inteligentes sobre otros planetas). Y si sucediese, sería una sorprendente demostración de una importante verdad conceptual: ni siquiera un amplio y complejo sistema de representaciones verbales y visuales tiene una conexión *intrínseca*, mágica, dada de una vez por todas, con lo que representa; una conexión independiente del modo en que fue causada y de lo que constituyen las disposiciones del sujeto hablante o pensante. Y esto es cierto tanto si el sistema de representaciones (palabras e imágenes, en nuestro ejemplo) está implementado físicamente —las palabras son palabras escritas o habladas y los dibujos son dibujos físicos— o tan sólo concebido mentalmente. Ni las palabras del pensamiento ni las imágenes mentales representan *intrínsecamente* aquello acerca de lo que tratan.

#### EL CASO DE LOS CEREBROS EN UNA CUBETA

He aquí una posibilidad de ciencia-ficción discutida por los filósofos: imaginemos que un ser humano (el lector puede imaginar que es él quien sufre el percance) ha sido sometido a una operación por un diabólico científico. El cerebro de tal persona (su cerebro, querido lector) ha sido extraído del cuerpo y colocado en una cubeta de nutrientes que lo mantiene vivo. Las terminaciones nerviosas han sido conectadas a una computadora supercientífica que provoca en esa persona la ilusión de que todo es perfectamente normal. Parece haber gente, objetos, cielo, etc.; pero en realidad todo lo que la persona (usted) está experimentando es resultado de impulsos electrónicos que se desplazan desde la computadora hasta las terminaciones nerviosas. La computadora es tan ingeniosa que si la persona intenta alzar su mano, el «*feedback*» que procede de la computadora le provocará que «vea» y «sienta» que su mano está alzándose. Por otra parte, mediante una simple modificación del programa, el diabólico científico puede provocar que la víctima «experimente» (o alucine) cualquier situación o entorno que él desee. También puede borrar la memoria de funcionamiento del cerebro, de modo que la víctima crea que siempre ha estado en ese entorno. La víctima puede creer incluso que está sentado, leyendo estas mismas palabras acerca de la suposición, divertida aunque bastante absurda, de que hay un diabólico científico que extrae cerebros de los cuerpos y los coloca en una cubeta de nutrientes que los mantiene vivos. Las terminaciones nerviosas se suponen conectadas a una computadora supercientífica que provoca en la persona la ilusión de ...

Cuando se menciona esta especie de posibilidad en una clase de Teoría del Conocimiento, el propósito no es otro que suscitar de un

modo moderno el clásico problema del escepticismo con respecto al mundo externo. (*¿Cómo podría usted saber que no se halla en esa situación?*) Pero esta situación es también un útil recurso para suscitar cuestiones en torno a la relación mente-mundo.

En lugar de imaginar un solo cerebro en una cubeta, podemos imaginar que los seres humanos (quizá todos los seres sintientes) son cerebros en una cubeta (o sistemas nerviosos en una cubeta, en el caso de algunos seres que sólo poseen un sistema nervioso mínimo, pero que ya cuentan como sintientes). Por supuesto, el diabólico científico tendría que estar fuera —¿o querría estarlo? Quizá no exista ningún diabólico científico, quizá (aunque esto es absurdo) el mundo consista en una maquinaria automática que está al cuidado de una cubeta repleta de cerebros y sistemas nerviosos.

Supongamos esta vez que la maquinaria automática está programada para ofrecernos a todos una alucinación *colectiva*, en lugar de unas cuantas alucinaciones separadas y sin relación. De forma que cuando me parece estar hablando con usted, a usted le parece estar oyendo mis palabras. Mis palabras no llegan realmente a sus oídos, por supuesto —porque usted no tiene oídos (reales), ni yo tengo boca o lengua reales. Pero cuando emito mis palabras, lo que ocurre en realidad es que los impulsos aferentes se desplazan desde mi cerebro hasta el ordenador, el cual a su vez provoca que yo «oiga» mi propia voz profiriendo esas palabras y «sienta» el movimiento de mi lengua, y que usted «oiga» mis palabras, y me «vea» hablando, etc. En este caso, nos comunicamos realmente, hasta cierto punto. Yo no estoy equivocado con respecto a su existencia real (sólo lo estoy con respecto a la existencia de su cuerpo y del «mundo externo», aparte de los cerebros). En cierta medida, tampoco importa que «el mundo entero» sea una alucinación colectiva; después de todo, cuando me dirijo a usted, usted oye realmente mis palabras, si bien el mecanismo no es el que suponemos. (Si fuéramos dos amantes haciendo el amor y no dos personas manteniendo una conversación, la insinuación de que únicamente somos dos cerebros en una cubeta podría ser molesta, desde luego.)

Deseo formular ahora una pregunta que parecerá obvia y bastante estúpida (al menos a algunos, incluyendo a algunos filósofos sumamente sofisticados), pero que tal vez nos sumerja con cierta rapidez en auténticas profundidades filosóficas. Supongamos que toda esta historia fuera realmente verdadera. Si fuéramos cerebros en una cubeta, ¿podríamos *decir* o *pensar* que lo somos?

Voy a argumentar en favor de la respuesta «no, no podríamos». En realidad, voy a argüir que la suposición de que realmente somos cerebros en una cubeta, pese a no violar ley física alguna y a ser perfectamente consistente con todas nuestras experiencias, no puede ser

verdadera. Y *no puede ser verdadera* porque, en cierto modo, se autorrefuta.

El argumento que seguidamente expondré es bastante inusual, y me llevó varios años convencerme de que era verdaderamente correcto. Pero es un argumento correcto. Lo que le da una apariencia tan extraña es su conexión con algunas de las más profundas cuestiones de la filosofía (se me ocurrió por primera vez cuando estaba estudiando un teorema de la lógica moderna, el teorema de Skolem-Löwenheim, y vi de repente una conexión entre este teorema y algunos argumentos de las *Investigaciones Filosóficas* de Wittgenstein).

Un «supuesto que se autorrefuta» es aquel cuya verdad implica su propia falsedad. Por ejemplo, consideremos la tesis de que *todos los enunciados generales son falsos*. Este es un enunciado general. De forma que si es verdadero, debe ser falso. Por lo tanto es falso. En ciertas ocasiones decimos que una tesis se «autorrefuta» si *la misma suposición de que la tesis es tomada en cuenta o enunciada* implica ya su falsedad. Por ejemplo, «no existo» se autorrefuta si *soy yo* (para cualquier «yo») quien lo pienso. De modo que uno puede estar seguro de que existe con sólo pensar en ello (como Descartes argumentó).

Demostraré que la suposición de que somos cerebros en una cubeta posee precisamente esta propiedad. Si podemos considerar su verdad o su falsedad, entonces no es verdadera (lo demostraré). Por lo tanto no es verdadera.

Antes de ofrecer el argumento, permítanme considerar el motivo por el que parece tan extraño que éste pueda siquiera ofrecerse (al menos para los filósofos que subscriben una concepción de la verdad-copia). Concedamos que es compatible con las leyes físicas que haya un «mundo posible» en el que todos los seres sintientes sean cerebros en una cubeta. (El discurso sobre este «mundo posible» suena como si hubiese un *lugar* donde cualquier suposición absurda fuese verdadera, y éste es el motivo por el que puede ser filosóficamente desorientador.) Los humanos de ese mundo posible tienen exactamente las mismas experiencias que tenemos *nosotros*. También piensan igual que nosotros (al menos por su mente pasan palabras, imágenes, formas de pensamiento). Aún así, estoy afirmando que podemos ofrecer un argumento que demuestre que no somos cerebros en una cubeta. ¿Cómo puede haberlo? ¿Por qué no podrían ofrecer tal argumento las personas que, en tal mundo posible, *son* realmente cerebros en una cubeta?

La respuesta será (básicamente) ésta: aunque esas personas pueden pensar y «decir» cualquier palabra que nosotros pensemos o digamos, no pueden «referirse» a lo que nosotros nos referimos. En particular, no pueden decir o pensar que son cerebros en una cubeta (*incluso pensando «somos cerebros en una cubeta»*).

## EL TEST DE TURING

Supongamos que alguien ha logrado inventar una computadora que pueda mantener realmente una conversación con un sujeto (o con tantas personas como podría hacerlo una persona inteligente) ¿Cómo podríamos decidir si la computadora es «consciente» o no?

El lógico británico A. Turing propuso el siguiente test<sup>2</sup>: permitamos que un sujeto mantenga una conversación con la computadora y otra conversación con una persona a quien no conoce. Si no puede decidir quién es la computadora y quién el ser humano, entonces (suponiendo que el test se repite en número suficiente de veces con distintos interlocutores), la computadora es consciente. En resumen, una máquina de computar es consciente si puede pasar el «Test de Turing» (las conversaciones no se mantienen cara a cara, por supuesto: el interlocutor no puede conocer la apariencia visual de sus dos contertulios; tampoco debe utilizarse la voz, ya que la voz mecánica podría sonar con un matiz distinto al de la voz humana). Imaginemos, más bien, que todas las conversaciones se mantienen con el soporte de una máquina de escribir eléctrica. El interlocutor mecanografía sus enunciados, preguntas, etc., y ambos contertulios —la máquina y el ser humano— responden por medio de un teclado eléctrico. (La máquina puede además *mentir*— si se le pregunta «¿Eres una máquina?»—, podría replicar «No, soy un asistente del laboratorio».)

Varios autores (de ningún modo hostiles a la idea de que una máquina pueda ser consciente) han criticado la idea de que este test sea realmente definitivo para la determinación de la conciencia. Pero nuestro tema no es éste. Deseo emplear la idea general del Test de Turing, la idea general de un *test dialógico de competencia*, para un propósito distinto: explorar la noción de *referencia*.

Imaginemos una situación en la que el problema no consiste en determinar si el contertulio es realmente una persona o una máquina, sino más bien determinar si usa referencialmente las palabras tal y como nosotros lo hacemos. El test más obvio consiste, de nuevo, en mantener una conversación y, si no surgen problemas, si el contertulio «aprueba» (en el sentido de que no puede ser distinguido de alguien de quien sabemos de antemano que habla el mismo lenguaje, se refiere a las mismas clases usuales de objetos, etc.), concluir que se refiere a los objetos tal y como nosotros lo hacemos. Cuando el propósito del Test de Turing sea precisamente el descrito, es decir, determinar la referencia compartida (*shared*) lo llamaré *Test de Turing para la Referencia*. Y del mismo modo que los filósofos han debatido la cues-

<sup>2</sup> A.M. TURING, «Computing Machines and Intelligence», *Mind*, 1950, reimpresso en A. R. Anderson (ed.), *Minds and Machines*.

tion de si el original Test de Turing constituye un test *definitivo* para la determinación de la conciencia, esto es, la cuestión de si una máquina que «apruebe» el test, no en una sola ocasión, sino regularmente, es *necesariamente* consciente, deseo discutir la cuestión de si el recién sugerido Test de Turing para la Referencia constituye un test definitivo para determinar la referencia compartida.

La respuesta resultará ser «NO». El Test de Turing para la Referencia no es definitivo. No hay duda de que es un test excelente en la práctica, pero no es lógicamente imposible (aunque sin duda es altamente improbable) que alguien pase el Test de Turing sin estar refiriéndose a nada. De aquí se sigue, como veremos, que es posible ampliar nuestra observación de que las palabras (y los textos completos y los discursos) no tienen una conexión necesaria con sus referentes. Aun cuando consideremos las reglas que deciden qué palabras pueden emitirse apropiadamente en ciertos contextos, en lugar de las palabras como tales —aun cuando consideremos *programas para el uso de las palabras*, según la jerga informática—, aún así, esas palabras no poseen una referencia determinada, a menos que esos programas se refieran por sí mismos a algo *extralingüístico*. Este será un paso crucial de cara a alcanzar la conclusión de que los cerebros en una cubeta no pueden referirse de ningún modo a nada externo (y, por consiguiente, no pueden decir *que* son cerebros en una cubeta).

Supongamos, por ejemplo, que me encuentro en la situación ideada por Turing (disputando el «juego de la imitación», según la terminología de Turing) y que mi contertulio es efectivamente una máquina. Supongamos que la máquina es capaz de ganar el juego («pasa» el test). Imaginemos que la máquina está programada para producir bellas respuestas en castellano en contestación a enunciados, preguntas, observaciones, etc., formulados en castellano, pero que no tiene órganos sensoriales (salvo la conexión con mi máquina de escribir eléctrica). (Turing no da por sentado que la posesión de otros órganos sensoriales o motores sea necesaria para la conciencia o la inteligencia, que yo sepa.) Supongamos que la máquina no sólo carece de ojos y oídos electrónicos, etc., sino que su programa, el programa que le permite llevar a cabo el juego de la imitación, también carece de dispositivos para la incorporación de *inputs* desde esos órganos, o para ejercer control sobre un cuerpo. ¿Qué tenemos que decir con respecto a esta máquina?

Me parece evidente que ni podemos ni debemos atribuir uso referencial a tal ingenio mecánico. Es cierto que la máquina puede platicar maravillosamente acerca del paisaje de Nueva Inglaterra, por ejemplo. Pero si tuviese enfrente una manzana, una montaña o una vaca, no podría reconocerlas.

Tenemos un artilugio para producir oraciones en respuesta a ora-

ciones. Pero estas oraciones no están conectadas con el mundo real. Si *acoplásemos* dos de estas máquinas y les dejásemos disputar el juego de la imitación entre sí, *continuarían «engañándose» una a otra eternamente, aun cuando el resto del mundo desapareciese*. No existen más razones para considerar que el discurso de la máquina acerca de manzanas se refiere a manzanas del mundo real, que para considerar que el «retrato» de la hormiga se refiere a Winston Churchill.

Lo que produce aquí la ilusión de referencia, significado, inteligencia, etc., es el hecho de que *tenemos* una convención de representación bajo la cual el discurso de la máquina se refiere a manzanas, campanarios, Nueva Inglaterra, etc. De modo parecido, y por la misma razón, existe la *ilusión* de que la hormiga ha caricaturizado a Churchill. Sólo que *nosotros* somos capaces de percibir, manipular y mantener trato con manzanas y campos. Existen «reglas de entrada al lenguaje» que nos conducen desde las experiencias con manzanas a preferencias tales como «Veo una manzana» y «reglas de salida del lenguaje» que nos llevan desde decisiones expresadas en forma lingüística («Voy a comprar algunas manzanas») a acciones distintas de la acción de hablar. Como la máquina carece de reglas de entrada o de salida del lenguaje, no hay ninguna razón para considerar el discurso de la máquina (o de ambas máquinas, en el caso que ideamos de dos máquinas jugando entre sí el juego de la imitación), como algo más que un juego sintáctico. Un juego sintáctico que *se parece* a un discurso inteligente, a buen seguro. Pero en la misma medida (y no en mayor) que la curva de la hormiga se parece a una mordaz caricatura.

En este último caso, podríamos haber argumentado que la hormiga habría dibujado la misma curva aunque Winston Churchill jamás hubiese existido. Pero no podemos elaborar un argumento que sea completamente paralelo para el caso de la máquina; si las manzanas, los árboles y los campanarios no hubieran existido, entonces, presumiblemente, los programadores no habrían dado a luz el mismo programa. A pesar de que la máquina no *percibe* manzanas, campos o campanarios, sus creadores-diseñadores sí lo hacen. Hay *cierta* conexión causal entre la máquina y las manzanas del mundo real, etc., por medio de la experiencia perceptiva y el conocimiento de sus creadores-diseñadores. Pero esta débil conexión difícilmente puede bastar para la referencia. No sólo es lógicamente posible —pese a ser fantásticamente improbable— que la misma máquina hubiera existido sin existir manzanas, campos y campanarios, etc. Aun cuando todas estas cosas *cesaran* de existir, la máquina seguiría platicando felizmente. Por este motivo, es absolutamente imposible sostener que la máquina es capaz de referirse a las cosas.

El punto relevante para nuestra discusión es que el Test de Turing no permite excluir a una máquina programada para no hacer otra cosa *salvo disputar* el Juego de la Imitación, y está claro que una má-

quina con tales características no se refiere a las cosas en mayor medida en que lo hace un tocadiscos.

## CEREBROS EN UNA CUBETA (DE NUEVO)

Permítasenos comparar los hipotéticos «cerebros en una cubeta» con las máquinas recién descritas. Obviamente, hay importantes diferencias. Los cerebros en una cubeta no tienen órganos sensoriales, aunque sí tienen elementos que funcionan como tales; es decir, hay terminaciones nerviosas aferentes e *inputs*, y estos *inputs* figuran en el programa de los cerebros en una cubeta en la misma medida en que figuran en el nuestro. Los cerebros en una cubeta son *cerebros*; por otra parte, son cerebros que *funcionan*, y lo hacen mediante las mismas reglas con las que funcionan los cerebros del mundo real. Por estas razones, parecería absurdo negarles conciencia o inteligencia. Pero el hecho de que sean conscientes o inteligentes no significa que sus palabras se refieran a lo que se refieren las nuestras. La cuestión que nos interesa es: cuando sus verbalizaciones contienen la palabra «árbol», ¿se refieren realmente a árboles? De forma más general: ¿acaso pueden referirse a objetos *externos*? (Como algo opuesto a los objetos que aparecen en la imagen producida por la maquinaria automática, por ejemplo.)

En orden a organizar nuestras ideas, permítasenos suponer que la máquina automática existe gracias a cierto tipo de azar o coincidencia cósmica (o quizá que siempre ha existido). En este hipotético mundo, se supone que la maquinaria automática no tiene un creador-diseñador inteligente. Como dijimos al comenzar este capítulo, podemos imaginar que todas las criaturas sensibles (aunque su sensibilidad sea mínima) se hallan en el interior de la cubeta.

Esta suposición no nos ayuda. Porque no hay conexión alguna entre la *palabra* «árbol» —tal como es utilizada por estos cerebros— y los árboles reales. Aunque no hubieran árboles reales, seguirían usando la palabra «árbol» como lo hacen, pensando los mismos pensamientos que piensan y teniendo exactamente las mismas imágenes que tienen. Sus imágenes, palabras, etc., son cualitativamente idénticas a las imágenes, las palabras, etc., que sí representan árboles en *nuestro* mundo; pero ya hemos visto (¡la hormiga otra vez!) que la semejanza cualitativa con algo que representa un objeto (Winston Churchill o un árbol) no hace que una cosa sea por sí misma una representación. En resumen, cuando los cerebros en una cubeta piensan «hay un árbol delante de mí» no están pensando en árboles reales, ya que no existe nada en virtud de lo cual su pensamiento «árbol» represente árboles reales.

Si esta conclusión parece precipitada, reflexionemos sobre lo si-

guiente: hemos visto que las palabras no se refieren necesariamente a árboles pese a estar ordenadas en una secuencia idéntica a un discurso que (si ocurriese en una de nuestras mentes) se referiría incuestionablemente a los árboles del mundo real. Tampoco «el programa», en el sentido de las reglas, prácticas y disposiciones de los cerebros a la conducta verbal, se refiere necesariamente a árboles, ni ocasiona la referencia a árboles a través de las conexiones que establece entre palabras y palabras, o entre señales de entrada *lingüística* y respuestas *lingüísticas*. Si estos cerebros representan, se refieren y piensan acerca de árboles (árboles reales, fuera de la cubeta), debe ser gracias al modo en que el programa conecta el sistema lingüístico con *inputs* y *outputs no-verbales*. Tales *inputs* existen efectivamente en el mundo de los cerebros en una cubeta (¿de nuevo aquellas terminaciones nerviosas aferentes y eferentes!) pero también veíamos que los «*sense-data*» producidos por la maquinaria automática no representan árboles (o algo externo) aun cuando se parezcan con exactitud a nuestras imágenes de los árboles. Así como una mancha de pintura podría parecer el dibujo de un árbol sin ser el dibujo de un árbol, veíamos que un «dato sensorial» podría ser cualitativamente idéntico a la «imagen de un árbol», sin ser la imagen de un árbol. En el caso de los cerebros en una cubeta, ¿cómo es que el hecho de que el lenguaje está conectado mediante el programa con *inputs* sensoriales que no representan intrínseca o extrínsecamente árboles (o cualquier otra cosa externa) puede posibilitar que todo el sistema de representaciones, el lenguaje-en-uso, represente o se refiera a árboles, o a cualquier otra cosa externa?

La respuesta es que no puede. Todo el sistema de *sense-data*, las señales motoras para las terminaciones eferentes y el pensamiento mediatizado verbal o conceptualmente y conectado mediante «reglas de entrada al lenguaje» con los *sense-data* (o con lo que sea) como *inputs* y mediante «reglas de salida del lenguaje» con las señales motoras como *outputs*, no tiene más conexiones con los árboles que las que la curva de la hormiga tenía con Winston Churchill. Una vez vemos que la semejanza cualitativa (ascendiéndola, si queremos, a identidad cualitativa) entre los pensamientos de los cerebros en una cubeta y los de alguien que exista en el mundo real no implica en modo alguno la mismidad de la referencia, no es difícil ver que no hay ninguna base para considerar que el cerebro en una cubeta se refiera a cosas externas.

#### LAS PREMISAS DEL ARGUMENTO

Ya he ofrecido el argumento prometido para demostrar que los cerebros en una cubeta no pueden pensar ni decir que son cerebros

en una cubeta. Queda sólo hacerlo explícito y examinar su estructura.

Según acabamos de afirmar, cuando el cerebro en una cubeta (en aquel mundo donde cada ser sintiente está y siempre estuvo dentro de una cubeta) piensa «Hay un árbol delante de mí», su pensamiento no se refiere a árboles reales. Según algunas teorías que seguidamente discutiremos, podría referirse a árboles que aparecen en la imagen, o a impulsos electrónicos que ocasionan experiencias de árboles, o a las características del programa que son responsables de esos impulsos electrónicos. Nuestras recientes afirmaciones no implican el rechazo de estas teorías, pues existe una estrecha conexión causal entre el uso de la palabra «árbol» en la cubeta-castellana y la presencia de árboles que aparecen en la imagen, la presencia de impulsos electrónicos de cierto tipo, y la presencia de ciertas características en el programa de la máquina. Según tales teorías, el cerebro *está en lo cierto*, *no se equivoca* al pensar «Hay un árbol delante de mí». Dada la referencia de «árbol» y de «delante de» en la cubeta-castellana, y suponiendo que una de estas teorías es correcta, entonces las condiciones de verdad de la oración «Hay un árbol delante de mí», cuando ocurre en una cubeta-castellana, son algo tan simple como el que haya un árbol aparente-en-la-imagen delante del mí en cuestión —en la imagen— o quizá que desde la maquinaria esté llegando el tipo de impulso electrónico que normalmente produce esta experiencia, o quizá que esté operando el dispositivo de la maquinaria que se supone que produce la experiencia de «árbol delante de mí». Y sin duda alguna, estas condiciones de verdad se satisfacen.

Por el mismo argumento, en la cubeta-castellana «cubeta» se refiere a cubetas-aparentes-en-la-imagen, o a algo relacionado (impulsos electrónicos o a características de programa), pero sin duda no se refiere a cubetas reales, ya que el uso de «cubeta» en la cubeta-castellana no tiene conexión causal con cubetas reales (aparte de la conexión que supone el hecho de que los cerebros en una cubeta no podrían usar la palabra «cubeta» si no estuvieran en presencia de una cubeta particular —la cubeta en que se hallan; pero en la cubeta-castellana esta conexión se obtiene entre el uso de toda palabra y esa cubeta particular: no es una conexión especial entre el uso de la palabra «cubeta» y las cubetas). De forma similar, en la cubeta-castellana, «fluido nutriente» se refiere un líquido-aparente-en-la-imagen o a algo relacionado (impulsos electrónicos o características del programa). Se sigue que si su «mundo posible» es el mundo real, y somos realmente cerebros en una cubeta, entonces lo que queremos decir con «Somos cerebros en una cubeta» es que *somos cerebros en una cubeta-aparente-en-la-imagen* o algo de esta índole (si es que queremos decir algo). Pero parte de la hipótesis de que somos cerebros en una cubeta

es que no somos cerebros en una cubeta-aparente-en-la-imagen (es decir, lo que estamos «alucinando» no es que somos cerebros en una cubeta). Así pues, si somos cerebros en una cubeta, entonces la oración «Somos cerebros en una cubeta» afirma algo falso (si es que afirma algo). En resumidas cuentas, si somos cerebros en una cubeta, entonces «Somos cerebros en una cubeta» es falso. Por lo tanto es (necesariamente) falso.

La suposición de que tal posibilidad tenga sentido surge de la combinación de dos errores: (1) tomar demasiado en serio la *posibilidad física*; y (2) manejar de forma inconsciente una teoría mágica de la referencia, es decir, una teoría en la que ciertas representaciones mentales se refieren necesariamente a ciertas cosas y clases de cosas externas.

Hay un «mundo físicamente posible» en el cual somos cerebros en una cubeta —¿qué significa esto, excepto que hay una descripción de tal estado de cosas que es compatible con las leyes de la física? Así como existe en nuestra cultura (desde el siglo XVIII) la tendencia a considerar la física como nuestra metafísica, es decir, a considerar las ciencias exactas como la descripción del «verdadero mobiliario último del universo» por tanto tiempo buscada, la consecuencia inmediata de esta tendencia es también cierta tendencia a considerar la «posibilidad física» como la mismísima piedra de toque de cómo podrían ser verdadera y realmente las cosas. La verdad es verdad física, la posibilidad, posibilidad física, y la necesidad, necesidad física, según esta opinión. Pero acabamos de ver, aunque por ahora sólo en el caso de un ejemplo muy artificial, que este punto de vista es erróneo. La existencia de un «mundo físicamente posible» en el cual somos cerebros en una cubeta (y siempre lo fuimos y siempre lo seremos) no significa que posible, verdadera y realmente podríamos *ser* cerebros en una cubeta. Lo que excluye esta posibilidad no es la física, sino la *filosofía*.

Algunos filósofos, anhelando hacer valer y al mismo tiempo minimizar las pretensiones de su profesión (típico estado de ánimo en la filosofía angloamericana del siglo XX) dirían: «Bravo. Ha mostrado que algunas cosas que parecen ser posibilidades físicas son en realidad imposibilidades *conceptuales*. ¿Qué tiene esto de sorprendente?»

Bien, no hay duda de que mi argumento puede describirse como un argumento conceptual. Pero cuando la actividad filosófica se describe como una búsqueda de verdades «conceptuales», suena a algo así como una *investigación acerca del significado de las palabras*. Y, no ha sido ésta nuestra tarea, desde luego.

Nuestra tarea ha consistido en examinar las *precondiciones del pensar acerca de algo, representar algo, referirse a algo*, etc. Y no hemos investigado estas precondiciones desentrañando el significado de es-

tas palabras y de estas frases (como podría hacer un lingüista, por ejemplo), sino *razonando a priori*. Y no en el antiguo sentido «absoluto» (ya que no pretendemos que las teorías mágicas de la referencia sean erróneas *a priori*), sino en el sentido de una investigación sobre lo que es *razonablemente* posible una vez *asumidas* ciertas premisas generales, o una vez *establecidas* ciertas suposiciones teóricas muy generales. Tal procedimiento no es «empírico», pero tampoco es completamente «a priori», aunque incorpora elementos de ambas formas de investigación. A pesar de ser falible y de depender de supuestos que podrían ser descritos como «empíricos» (por ejemplo, el supuesto de que la mente no tiene otro acceso a las cosas externas o a las propiedades que el que le proporcionan los sentidos), mi procedimiento se halla en estrecha relación con lo que Kant llamó «investigación trascendental», ya que consiste, repito, en una investigación sobre las *precondiciones* de la referencia y, por ende, del pensamiento —precondiciones que se encuentran incorporadas en la naturaleza de nuestras mentes, aunque no son (como Kant creía) completamente independientes de suposiciones empíricas.

Una de las premisas del argumento es obvia: las teorías mágicas de la referencia son erróneas, y no sólo erróneas con respecto a las representaciones físicas, sino también en lo que concierne a las mentales. La otra premisa establece la imposibilidad de referirnos a ciertos tipos de cosas, por ejemplo, *a árboles*, sin haber tenido interacción causal con ellas<sup>3</sup>, o, en su caso, con otras cosas en cuyos términos puedan describirse las primeras. Pero ¿por qué hemos de aceptar estas premisas? Puesto que constituyen el marco general en el que opera mi argumentación, es el momento de examinarlas más detenidamente.

#### RAZONES PARA NEGAR QUE HAYAN CONEXIONES NECESARIAS ENTRE LAS REPRESENTACIONES Y SUS REFERENTES

Anteriormente apuntaba que algunos filósofos (el más célebre, Brentano) han adscrito una facultad a la mente, la «intencionalidad», que la capacita precisamente para *referirse* a las cosas. Evidentemente, descarto que esto constituya una solución. Pero ¿con qué derecho lo hago? ¿He ido demasiado deprisa?

Estos filósofos no afirmaban que fuera posible pensar en cosas

<sup>3</sup> Si los cerebros en una cubeta fueran a tener conexión causal con árboles en el futuro, por ejemplo, entonces quizá podrían referirse *ahora* a árboles mediante la descripción «Las cosas a las que me referiré como «árboles» en tal y cual ocasión futura». Pero, en nuestro caso, los cerebros en una cubeta no se libran *nunca* de la cubeta, y por *tanto* no entran *nunca* en conexión causal con árboles, etc.

externas o en propiedades sin usar ningún tipo de representación. Y aceptarían el argumento que ofrecí anteriormente y que comparaba los *sense data* con el «dibujo» de la hormiga (el argumento que se servía de la historia de ciencia ficción acerca del «dibujo» de un árbol producido por una mancha de pintura, que daba lugar a *sense data* cualitativamente similares a nuestras imágenes visuales de los árboles, pero que no estaban acompañados por ningún *concepto* de árbol), como muestra de que las *imágenes* no se refieren a las cosas de un modo necesario. Si existen representaciones mentales que se refieran necesariamente (a cosas externas), deben ser de la naturaleza de los *conceptos*, no de las imágenes. Pero ¿qué son los *conceptos*?

Cuando practicamos la introspección no percibimos «conceptos» fluyendo por nuestra mente así como así. Dondequiera y cuandoquiera que detengamos el flujo de pensamiento, lo que atrapamos son palabras, imágenes, sensaciones, sentimientos. Cuando digo en voz alta mis pensamientos, no los pienso por segunda vez. Oigo mis palabras como usted la oye. No hay duda de que cuando creo las palabras que profiero me siento en un estado «diferente» a cuando no las creo (pero a veces, cuando estoy nervioso, o ante una audiencia hostil, me parece estar mintiendo, pese a saber que estoy diciendo la verdad); también me siento en un estado distinto cuando profiero palabras que comprendo a cuando profiero palabras que no comprendo. Pero puedo imaginar sin dificultad a alguien que piense estas palabras (en el sentido de repetir las mentalmente) y experimente el mismo sentimiento de comprensión, afirmación, etc., que yo tengo, pero percatándose (o siendo despertado por un hipnotizador) un minuto más tarde de que no entendía lo que acababa de pasar por su mente, y de que ni siquiera entendía el lenguaje en el que estaban esas palabras. No pretendo que esto sea muy probable, sino que es perfectamente imaginable. Y esto no muestra que los *conceptos* son palabras (o imágenes, sensaciones, etc.) sino que atribuir a alguien un concepto o un pensamiento es algo completamente distinto de atribuirle alguna «representación» mental, alguna entidad o evento susceptible de introspección. La razón decisiva para sostener que los *conceptos* no sean representaciones mentales que se refieran intrínsecamente a las cosas es que ni tan siquiera son representaciones mentales. Los *conceptos* son símbolos que se usan de cierto modo; los símbolos pueden ser públicos o privados, entidades mentales o físicas, pero aun cuando los símbolos sean «mentales» y «privados», el propio símbolo, con independencia de su uso, no es el concepto. Y los símbolos no se refieren de por sí intrínsecamente a ninguna cosa.

Podemos observar esto realizando un experimento mental muy simple. Supongamos que ni usted ni yo podemos distinguir un olmo de

un haya. Aún así, afirmamos que la referencia de «olmo» en mi discurso es idéntica a la referencia de «olmo» en el discurso de cualquier otra persona, v.gr. olmos, y que la extensión de «haya» es el conjunto de todas las hayas (es decir, el conjunto de cosas de las que la palabra «haya» se predica con verdad) tanto en su discurso como en el mío. ¿Es posible creer que la diferencia entre la referencia de «olmo» y la referencia de «haya» descansa en una diferencia en nuestros *conceptos*? Mi concepto de olmo es exactamente el mismo que mi concepto de haya (me ruboriza confesarlo). (Esto demuestra, de paso, que la determinación de la referencia es social y no individual; tanto usted como yo confiamos en los expertos que *si pueden* distinguir los olmos de las hayas.) Si alguien intenta mantener heroicamente que la diferencia entre la referencia de «olmo» y la referencia de «haya» en *mi* discurso se explica por una diferencia en mi estado psicológico, dejémosle imaginar una Tierra Gemela en la que las palabras están trocadas. La Tierra Gemela es muy parecida a la Tierra; en realidad, aparte del hecho de que las palabras «olmo» y «haya» están intercambiadas, el lector puede suponer que la Tierra Gemela es exactamente como la Tierra. Supongamos que tengo un *doble* en la Tierra Gemela que es idéntico a mí, molécula por molécula, en el sentido en que dos corbatas pueden ser *idénticas*. Si usted es dualista, suponga que mi *doble* tiene los mismos pensamientos verbalizados que yo, experimenta los mismos datos sensoriales, posee las mismas disposiciones, etc. Es absurdo pensar que su estado psicológico es diferente del mío: pero aun así, su palabra «olmo» representa *hayas*, y mi palabra «olmo» representa olmos. (De forma similar, si el «agua» de la Tierra Gemela fuese un líquido diferente —XYZ y no H<sub>2</sub>O, por ejemplo— entonces «agua» representaría líquidos distintos cuando se usa en la Tierra Gemela o cuando se usa en la Tierra.) Contrariamente a lo que afirma la doctrina que se ha mantenido entre nosotros desde el siglo diecisiete, *los significados no están en la cabeza*.

Hemos visto que poseer un concepto no consiste en poseer imágenes (de árboles, por ejemplo— o incluso de imágenes «visuales» o «acústicas» de oraciones, o de discursos enteros, dicho sea de paso), ya que alguien podría poseer cualquier sistema de imágenes que a usted se le antoje sin poseer por ello la *capacidad* de utilizar las sentencias de modos situacionalmente adecuados, considerando los factores lingüísticos —como afirmamos con anterioridad— y también a los no lingüísticos, como los determinantes de la «adecuación situacional». Un hombre podría tener todas las imágenes que usted quiera, y, sin embargo, no saber qué ha de hacer cuando se le dice «Señale el árbol», aun cuando haya presente un grupo de árboles. Hasta podría tener la imagen de lo que tiene que hacer, y, con todo, no saber lo que tiene que hacer. Pues la imagen, si no está acompañada por la capacidad de actuar de cierta manera, es sólo *imagen*, y la capaci-

dad de actuar de acuerdo «con un dibujo» es algo que se puede tener o no. Ese hombre podría representarse a sí mismo señalando un árbol, solamente con el ánimo de considerar algo lógicamente posible, después de que alguien produjese la secuencia de sonidos —que para él no tienen significado— «Por favor, señale un árbol». Y a pesar de todo no sabría que debe señalar un árbol, ni comprendería la exhortación «Señale un árbol».

He establecido que el criterio para poseer un «concepto efectivo» es la capacidad de usar ciertas oraciones, aunque este criterio puede ser liberalizado fácilmente. Podríamos permitir que los elementos del simbolismo no fuesen palabras del lenguaje natural, por ejemplo, y también podríamos dar entrada a fenómenos mentales como las imágenes y otras clases de eventos internos. Lo esencial es que éstos tengan la misma complejidad, la misma capacidad de ser combinados entre sí, etc., que las oraciones del lenguaje natural. Porque aunque una determinada representación —por ejemplo, un destello azul— podría servirle a un determinado matemático como expresión interior de toda la prueba del Teorema de los Números Primos, no tendríamos la tentación de afirmarlo (y sería falso hacerlo) si el matemático no pudiese desmembrar ese «destello azul» en pasos separados y conexiones lógicas. Pero, sin tener en cuenta cuáles son los tipos de fenómenos internos que permitimos como posibles *expresiones* del pensamiento, podrían ofrecerse argumentos exactamente similares al anterior que demostrarían que el entendimiento no está constituido por fenómenos, sino más bien por la capacidad del sujeto pensante para *emplear* y para producir fenómenos adecuados en las circunstancias adecuadas.

Acabo de ofrecer una versión muy abreviada del argumento de Wittgenstein de las *Investigaciones Filosóficas*. Si es correcto, entonces la pretensión de comprender el pensamiento a la luz de la denominada «investigación fenomenológica» se halla fundamentalmente desencaminada; pues lo que los fenomenólogos no ven es que lo que están describiendo es la *expresión interna* del pensamiento, pero la comprensión de esta expresión —la *comprensión* de los *pensamientos de uno mismo*— no es algo que acaezca, sino el resultado de una *capacidad*. Nuestro ejemplo de un hombre que simula pensar en japonés (y engaña a un telépata japonés) muestra de inmediato la falibilidad de una aproximación fenomenológica al problema de la *comprensión*. Pues aún habiendo alguna cualidad susceptible de introspección que estuviese presente cuando, y sólo cuando, uno *comprendiese realmente* (y esto, de hecho, parece ser falso con respecto a la introspección), esta cualidad estaría tan sólo *correlacionada* con la comprensión, y aún así sería posible que el hombre que engaña al telépata japonés se hallase en posesión de esa cualidad y, a pesar de todo, no entendiese ni una palabra de japonés.

Por otra parte, consideremos a un hombre (perfectamente posible) que no tiene ningún «monólogo interior». Habla perfectamente el castellano, y si le preguntamos cuáles son sus opiniones con respecto a un asunto dado, nos las ofrecerá detenidamente. Mas nunca piensa (con palabras, imágenes, etc.) cuando está hablando en voz alta; ni «pasa nada por su cabeza», salvo que (por supuesto) oye su propia voz hablando, tiene las impresiones sensoriales usuales de sus alrededores, más un «sentimiento general de comprensión». (Quizá tiene el hábito de hablar a solas.) Cuando mecanografía una carta, o cuando va de compras, etc., no experimenta un «flujo de pensamiento» pero sus acciones son inteligentes y propositivas, y si alguien le aborda y le pregunta «¿Qué está haciendo usted?», ofrecerá una respuesta perfectamente coherente.

Este hombre parece perfectamente imaginable. Nadie vacilaría en decir que es consciente, que no le gusta el rock and roll (si expresase con frecuencia una fuerte aversión por el rock and roll), etc., solamente porque no piensa conscientemente excepto cuando habla en voz alta.

De todo esto se sigue que (a) ningún conjunto de eventos mentales —imágenes o acontecimientos y cualidades más abstractas— *constituye* por sí mismo la comprensión; y (b) ningún conjunto de eventos mentales es *necesario* para la comprensión. En particular, *los conceptos no pueden ser idénticos a objetos mentales de ningún tipo*. Pues, aun suponiendo que entendemos por evento mental algo susceptible de introspección, cualquiera que sea éste, acabamos de ver que podría estar ausente en un hombre que entiende la palabra apropiada (y por tanto posee el concepto efectivo) y presente en un hombre que de ningún modo posee el concepto.

Volviendo ya a nuestra crítica de las teorías mágicas de la referencia (tópico que también interesaba a Wittgenstein), vemos que, por una parte, estos «objetos mentales» que *podemos* detectar introspectivamente —palabras, sentimientos, imágenes, etc.— no se refieren a algo intrínsecamente en mayor medida que lo hace el dibujo de la hormiga (y por las mismas razones), mientras que los intentos de postular objetos mentales especiales, «conceptos», los cuales *sí* tienen una conexión necesaria con sus referentes —y que sólo los fenomenólogos entrenados pueden detectar— *incurren en un error lógico*, porque los conceptos son (al menos en parte) *capacidades*, y no cosas que acontecen en la mente. La doctrina que defiende la existencia de representaciones mentales que se refieren necesariamente a las cosas externas no es sólo mala ciencia natural; es también mala fenomenología y confusión conceptual.